DOCUMENT RESUME

ED 066 928 | 24 | EM 010 329

AUTHOR	Yens, David P.
TITLE	The Interaction Between Reward Preference and Task Difficulty in a Computer Assisted Instructional Setting. Final Report.
INSTITUTION	Pennsylvania State Univ., University Park.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
REPORT NO	P-1-C-054
BUREAU NO	BR-1-C-054
PUB DATE	Nov 71
GRANT	OEG-3-71-0132
NOTE	253p.
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	Arithmetic; *Complexity Level; *Computer Assisted Instruction; Elementary School Mathematics; Elementary School Students; *Motivation; Positive Reinforcement; *Reinforcement; Rewards; *Task Performance

ABSTRACT

Two aspects of motivation were investigated in this study: the use of individual incentives to enhance learning, and the effect of different levels of task difficulty on the effectiveness of these individualized incentives. One hundred-forty-two fourth and fifth grade students were subjects. Individual preferences for rewards that would serve as incentives were first assessed through a Reward Preference Inventory, and then each subject was randomly assigned to one of three levels of incentive (most preferred reward, least preferred reward, or no reward) and to one of four levels of task difficulty in a computer-assisted arithmetic lesson. Task time and number of problems attempted were dependent variables. While results were generally not congruent with expectations, there was some indication that the promise of a reward will cause students to work longer on a task than no promise of a reward, and the promise of a most preferred reward will be more potent than the promise of a reward of lesser preference. In addition, no support was found for any of the theories of interaction between motivation and task difficulty. (Author/SH)

ERIC

1-2 25-4
PA 24
VC
EM

ED 066923

U. S. DEPARTMENT OF
HEALTH, EDUCATION AND WELFARE

Office of Education

National Center for
Educational Research
and Development

(Regional Research Program)

Final Report

Project No. 1-C-054

Grant No. OEG-3-71-0132

The Interaction Between Reward Preference and Task Difficulty in a Computer Assisted Instructional Setting

David P. Yens

Computer Assisted
Instruction Laboratory

The Pennsylvania
State University

University Park,
Pennsylvania

November, 1971

FILMED FROM BEST AVAILABLE COPY

EM 010 329

ED 066923

FINAL REPORT

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Project No. 1-C-054

Grant No. OEG-3-71-0132

THE INTERACTION BETWEEN REWARD PREFERENCE AND TASK DIFFICULTY IN A COMPUTER ASSISTED INSTRUCTION SETTING

David P. Yens

The Pennsylvania State University

University Park, Pennsylvania

November, 1971

The research herein was performed pursuant to a Grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development

(Regional Research Program)

1 A

ABSTRACT

This study investigated two aspects of motivation: the use of individual incentives to enhance learning of children and the effect of different levels of task difficulty on the effectiveness of these individualized incentives. It was also designed to provide evidence relating to several theoretical approaches to the interaction of incentive with task difficulty.

Fourth and fifth grade students worked arithmetic problems presented via computer terminals with the expectation that if they did well they would receive a reward of high or low rated preference or no reward at all. Four percentage ranges of difficulty were used with the problems; students could be assigned to work any of 23 kinds of problems graded in difficulty. The computer insured that performance was maintained within specified limits. Measures of persistence were time on the task and the number of problems attempted.

Limited indications of the differential motivational value of incentive preferences were found. The expected finding of the interaction was not supported; there appears to be no interaction between incentive and task difficulty for the variables studied. No evidence was found that would support the theories investigated.

ERIC

ACKNOWLEDGEMENTS

A research project of this type requires the help and encouragement of many people. The primary debt of gratitude is to Dr. Harold E. Mitzel, under whose direction this project was conducted, for his valuable guidance, suggestions, encouragement, and patience. The suggestions advice, and criticisms of Dr. G. Phillip Cartwright, Dr. Robert Seibel, and Dr. Paul Weener were valuable contributions to the study.

Several of the staff of the Computer Assisted Instruction Laboratory at the Pennsylvania State University were very helpful. The support and encouragement of Dr. Keith A. Hall are especially appreciated. Debts of gratitude are also due to Dr. Karl G. Borman, Mr. Terry A. Bahn, and Mr. William Matson for their technical assistance.

Special thanks are due to the administrators and teachers of the schools that furnished students for the study. Pretesting of some materials was conducted at the Park Forest Elementary School in State College, Pennsylvania. The study was carried out at the Athens Main Elementary School in Athens, Pennsylvania. Without the assistance of the people in these schools the experiment could not have been conducted.

Finally, profound gratitude is due to the author's wife, Alice, who provided valuable assistance during several phases of the project and whose constant encouragement culminated in a successful conclusion.

TABLE OF CONTENTS

LIST OF TABLES

ERIC

ERIC

LIST OF FIGURES

CHAPTER I

PROBLEM AND OBJECTIVES

Introduction and Problem

The motivation of students has been a major concern of professional personnel in all fields and levels of education. The question of how to motivate students has been a perennial one. Although a great deal of study has been devoted to the motivation of human behavior, little of this research has had an impact on the motivation of students in the classroom or in other educational settings, and very little seems to be applicable. Cartwright (1970) clearly points to a need for studying the motivational processes used by teachers. Although teachers use some general motivational procedures to establish incentive, maintain interest and reinforce learning behaviors,these procedures are usually applied indiscriminately. Furthermore, the most often used reward is praise or adult approval, but there is little evidence that this is an effective reinforcer for all children. Research (as well as common sense) has indicated that there is considerable variability among children as to preference for rewards (see Cartwright, 1970, p. 152).

An educational process such as computer assisted instruction (CAI) can provide a controlled environment for the study of the effect of many motivational mechanisms on a variety of educational processes. This environment can be made more "classroom-like" than most laboratories, permitting investigations to be conducted in a realistic

ERIC

educational setting. Thus, the results of such studies should have applicability to educational processes other than CAI, including classroom teaching.

At present, it has been found that work with the computer may in itself be motivating for most students (Mitzel, Hall, Suydam, Jansson, and Igo, 1971). However, there are some students who may need an added incentive for learning or for performing. Some evidence of this need has already emerged from two studies connected with a long-range evaluation of a mathematics curriculum that makes extensive use of CAI. Algebra and general mathematics students were taught predominantly by CAI in two Pennsylvania secondary schools. An evaluation of the effectiveness of CAI (Mitzel, et. al., 1971) found a decrease in favorable attitude toward CAI throughout the year in all classes and an initial increase in favorable attitude toward mathematics followed by a decrease near the end of the year. The attitude toward mathematics of a cohort non-CAI group, although typically lower than that of the CAI group, showed a small but consistent gain throughout the year. The CAI classes were also subject to high absence rates in one school which may have resulted in lower mastery levels than might have resulted from a lower absence rate. It may be speculated that learning and motivation might have been affected by the attitudes toward CAI and mathematics such that consistently improving attitudes might have led to better learning than that obtained (which, in fact, was typically better than than of the cohort group on measures related to the objectives of the CAI course). Confer (1971) found that the holding power of CAI was not any greater than than for conventional instruction; several students taking a remedial summer course in mathematics via CAI dropped out of the course.

Research on the use of contingency management for students with learning or behavior problems has indicated that specific incentives (candy, free play time, etc.) can be highly effective in producing desirable behaviors in children (e.g., Ulrich, Louisell, and Wolfe, 1971; O'Leary and Drabman, 1971). However, the administration of these rewards requires the attention of a teacher or aid. The computer could keep the student informed of his progress toward attaining an incentive and, upon request, provide the teacher with a listing of the rewards due each student.

Cartwright (1970) has found some evidence that task performance by elementary school children is related to their preference for the reward they receive upon completion of the task. Hunt (1961, 1965, 1971) has proposed that the magnitude of motivation required for optimum learning or performance depends upon the difficulty of the learning task. Furthermore, an interaction between the difficulty of the task and the strength of motivational variables has been found by several investigators (Yerkes and Dodson, 1908; Spence, 1958; Broadhurst, 1959; Atkinson and Litwin, 1960; O'Neil, Speilberger and Hansen, 1969). Unfortunately, precision has been lacking in the specification of the difficulty of a task, a precision that can be gained through the application of computer control. As a result of studies of these variables, several theories and approaches have been developed that are concerned with the interaction of motivation with task difficulty. However, these approaches result in conflicting predictions for the effect of the interaction variables upon task performance.

Individualized adjustment of task difficulty can be only roughly accomplished by researchers and classroom teachers. A computer could be of aid in that it could be used to establish and maintain levels of task difficulty. Furthermore the increasing use of computer assisted instruction in school classrooms makes the future control of drill-and-practice and tutorial learning task difficulty not only possible but perhaps desirable. A computer in the classroom would be able to monitor the performance of each student and also keep a record of rewards. Clearly, if educational practitioners are to make effective use of the computer, more needs to be known about the impact of different levels of subject matter difficulty on students and about its relationship to motivational variables.

If computer assisted instruction is to become a truly viable educational tool, research is needed to determine the optimum program and environmental characteristics for learning. Two of these characteristics have been identified, motivation (which is central to all learning) and task difficulty. The study of the interaction between these two variables has major theoretical implications and also provides a powerful means of evaluating the effects of the variables individually as well as the effect of one upon the other.

The selection of the specific variables for study was based upon the educational significance and generalizability of the variables, applicability to all competing theoretical formulations, and ease of manipulation. These criteria led to the use in this study of incentive as the motivational variable, several kinds of arithmetic problems as the task variable, and two different measures of persistence as the dependent variable. The theories and criteria will be treated in detail in the next chapter.

Specifically, two problems were considered. The first was the evaluation of competing theories of the motivation by task difficulty interaction by using independent variables that afford a common denominator for all the theories, e.g., variables that are not a part of the theoretical formulations. The second problem was delineating the effects of manipulating various categories of rewards on the performance of elementary school children and evaluating the effects of different levels of task difficulty in mathematics drill and practice administered by computer assisted instruction.

Purposes

Several outcomes were anticipated for this research.

1. It will provide evidence relevant to several theories concerning the interaction of task difficulty with preference for reward with respect to persistence of problem solving behavior.

2. It would provide further information on the effectiveness of measuring the preference for incentive through the medium of computer assisted instruction.

3. It would evaluate the effectiveness of individualizing rewards for elementary school children.

If differences were found in the motivational value of incentives and in performance at different levels of task difficulty, then the project would also:

4. Provide a basis for developing individualized incentive contingencies for students learning via computer assisted instruction, and

5. Provide information concerning the specification of optimum individualized levels of difficulty for drill and practice tasks.

Terminology

Many of the terms used in this report, such as motivation, drive, reinforcement, incentive, reward, and difficulty have been defined in different ways. The usage by the acknowledged authors is assumed for the review of the literature. A definition of terms is provided in Appendix A as well as a list of abbreviations used.

ERIC

CHAPTER II

RELATED RESEARCH

The Yerkes-Dodson Law and two theories of motivation have generated a great deal of research into the interaction of motivation with task difficulty. A third theory, that of Hunt, appears not to have been tested with elementary school-aged children. The present research may be the first to incorporate into the design a test of Hunt's hypotheses of intrinsic motivation.

This review will be concerned initially with the theories and research concerning the interaction of motivational variables with task difficulty. The last section will deal with research relevant to the variables of specific interest in this study.

Studies of the Interaction of Motivation with Task Difficulty

Yerkes-Dodson Law. In a 1908 paper, Yerkes and Dodson reported evidence of a curvilinear relationship between strength of motivation and learning in mice. This relationship was in the form of an inverted U-shaped curve, which indicated that maximum learning performance occurred at some motivational level intermediate between the minimum and maximum levels applied. (Hebb [1955] has postulated a similar curve describing the relationship between the level of cortical arousal and the level of cue function in which there is an optimal level of arousal for effective behavior.) In addition, Yerkes and Dodson noted an interaction between motivation and task difficulty in which motivation for a learning task decreased with an increase in task difficulty. For a more difficult task, the optimum learning occurred at a lower shock intensity than for a less difficult task. This interaction has been called the Yerkes-Dodson Law by Broadhurst (1959) and others. The obtained relationships are shown in Figure 1.

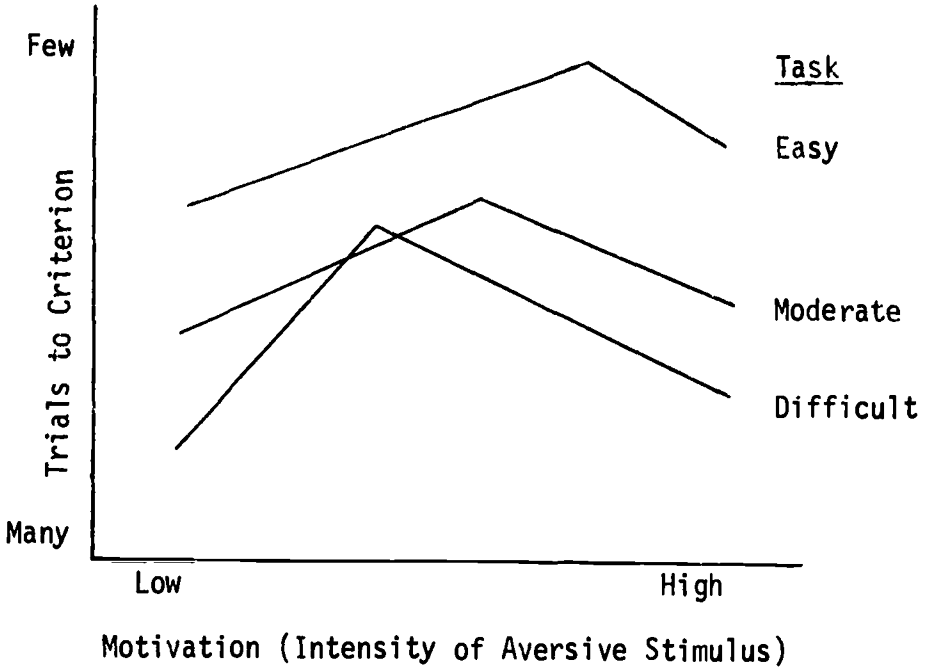


Fig. 1. Typical relationship found between motivation and task difficulty due to the Yerkes-Dodson Law. Based on Broadhurst (1959), p. 322.

The most consistently reliable indications of a curvilinear interaction between motivation and task difficulty are found in studies with animals in support of the Yerkes-Dodson Law. In their original study, Yerkes and Dodson (1908) used food reward for correct choices by mice and shock for incorrect ones in a brightness discrimination task. With a simple discrimination, faster learning occurred with increases in shock level until an optimum intensity was reached, then at shock levels above this optimum slower learning occurred (Figure 1). This curvilinear relationship has been repeatedly verified. However, with the addition of a task difficulty variable, discrimination difficulty, it was found that the optimum shock intensity was found at lower levels for more difficult discriminations.

It should be observed that the Yerkes-Dodson Law refers to an interaction of motivation with task difficulty and assumes a curvilinear relationship between motivation and performance. Thus, evidence of the curvilinear relationship and the interaction is necessary for support of the law.

Direct evidence supporting the Yerkes-Dodson Law has been provided by Dodson (1915), Cole (1911), Broadhurst (1959), and Hammes (1956). All evidence that has bearing on the theory is listed in Table 1. Dodson (1915) used kittens in a brightness discrimination task with shock for wrong responses. Cole (1911) used electric shock as a negative reinforcer and access to a warmed chamber for reward of chicks in a similar kind of problem. Hammes (1956) used three levels of shock and two levels of discrimination difficulty (black and white areas) with rats. More recently, Broadhurst (1959) used rats in an underwater brightness discrimination task. Motivation was provided by depriving the rats of air (keeping them submerged for various periods of time before release into the apparatus). His results were quite similar to those of Yerkes and Dodson; data analysis produced significant main effects and a significant difficulty by motivation interaction. Broadhurst discussed the application of the Yerkes-Dodson Law to human psychology, with particular reference to the drive properties of anxiety and to abnormal psychology.

Table 1

Studies of Motivation and Task Difficulty
Related to the Yerkes-Dodson Law

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Yerkes & Dodson (1908) Brightness descrimination by mice	Level of shock Food reward Discrimination difficulty	Trials to criterion	Complex curvilinear relationship between motivation and performance; apparent interaction between motivation and discrimination difficulty	Supports Yerkes-Dodson Law
Dodson (1915) Brightness discrimination by kittens	Replication of Yerkes-Dodson experiment with similar results.			
Cole (1911)	Similar study using chicks; similar results			
Broadhurst (1959) Underwater brightness discrimination - rats	Period of air deprivation Discrimination difficulty	Mean correct trials per 100 trials	Results similar to those of Yerkes and Dodson.	

Table 1 (Continued)

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Hammes (1956) Hue discrimination by rats	Level of shock Discrimination difficulty	Trails to criterion Mean errors	Results similar to those of Yerkes and Dodson	Can be held to support Yerkes-Dodson, but interpreted in terms of competing response tendencies - possible support for Spence-Taylor
Willett (1964a) Serial reaction task (moderate complexity)	Drive level (Eysenck - refers to drive variable used in series of studies in Eysenck, 1964)	Correct response - -	Lo drive>Hi drive ($p<.01$)	This series of studies was held to support the hypothesis of curvilinear relationship between drive and performance (Eysenck 1964)
		Error response - - -	Hi drive>Lo drive ($p<.01$)	
		All responses - - -	No difference (absolute differences very small, however)	

Table 1 (Continued)

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Eysenck & Gillan (1964). Mirror drawing (easy-moderate diff.)	Drive level (Eysenck)	Time for completion - Accuracy - - - - - -	Hi drive>Lo drive Hi drive>Lo drive	They also support the Spence-Taylor Theory
Eysenck & Warwick (1964) Multiple-choice reaction task (complex)	Drive level (Eysenck)	Mean correct responses (180 trials)	Lo drive>Hi drive ($p<.001$)	
Stennett (1957) Visual tracking	Incentive (4 levels)	EMG gradients	Curvilinear relationship between incentive and task performance	Partial support for Yerkes-Dodson
Wickens (1942) Arithmetic	Problem difficulty Competitive *vs* isolation	performance	Curvilinear relationship between difficulty and performance; interaction between difficulty and motivation indicated	Supports Yerkes-Dodson

ERIC

Table 1 (Continued)

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Spence (1958) Series of studies with eyeblink	Anxiety level Intensity of air puff (UCS)	Percent correct responses	Curvilinear relationship between intensity and percent correct responses for all levels of anxiety	Partial support for Yerkes-Dodson; support for Spence-Taylor
Fang (1966) Concept formation	Concept difficulty Incentive	Correct responses	No difference between incentive groups; no interaction Possible experimental weaknesses	No support for Yerkes-Dodson or Spence-Taylor
Fantino, Kasdon, and Stringer (1970) Detour problem	Percent normal body weight Degree of detour wraparound	Speed of solution	Monotonic increase, significant main effects	No support for Yerkes-Dodson

A recent test of the Yerkes-Dodson Law using hunger motivation of pigeons failed to confirm the findings of studies using aversive motivation. Fantino, Kasdon, and Stringer (1970) used a detour problem with three degrees of difficulty. A cup of food was placed behind a wire screen. Three levels of motivation were provided through food deprivation. Significant differences were found for the main effects of difficulty and deprivation level, but the slope of each difficulty level was linear and monotonic.

Several studies have been done with human subjects that have bearing upon the Yerkes-Dodson Law. The studies by Eysenck and his co-workers (Eysenck, 1964) are noteworthy for their operational control over the motivation variable. They were able to perform a series of experiments in an industrial setting. Their high-drive group consisted of candidates taking tests for a limited number of positions in a highly desirable apprentice training school. The tasks were administered as part of the entrance test battery. The low-drive group consisted of students already attending the school; the task was represented as an experiment with no reflection upon their standing in school. Evidence of a significant difference in drive was obtained through psychogalvanometric evaluation.

As a result of an analysis of a series of studies of different degrees of complexity, Eysenck (1964) pointed to evidence of a curvilinear relationship between drive and performance. In one study (Eysenck and Warwick, 1964) a mirror drawing task of easy to moderate difficulty resulted in superiority of the high drive subjects on the time and accuracy of the response. Willett (1964a) found no difference in the total number of responses by high and low drive subjects on a self-paced five choice serial reaction task of moderate complexity but found reliable differences between the groups on mean correct responses (low drive groups superior), and mean error responses (high drive group made more than the low drive group). However, the magnitude of the differences was small. On an externally paced multiple choice reaction task of high complexity (Eysenck and Gillan, 1964), the low drive subjects performed significantly better than high drive subjects. These studies appear to provide support for the Yerkes-Dodson Law in that the predicted curvilinear relationship between motivation and performance and the motivation--task difficulty interaction are indicated.

Willett (1964b) obtained contrary evidence with a paired-associate learning task using two lists of different difficulty. For both lists, the high drive subjects responded significantly faster than the low drive subjects. No interaction was found between drive and list difficulty.

Three studies of the motivation-by-task difficulty interaction have used incentives that might be acceptable in schools, although the results are mixed. Stennett (1957) used four levels of incentive, consisting of encouragement and money, with a visual tracking test. He measured EMG gradients to provide an indication of arousal and found that incentives of increasing value increased the steepness of the gradients. This was held by to be a demonstration of a curvilinear relationship between motivation and task performance. The most efficient tracking was found to be associated with an intermediate slope of the EMG gradients, which provides some support for the Yerkes-Dodson Law.

Wickens (1942) reported a study employing four levels of difficulty of arithmetic tasks in competitive or isolated conditions. Apparently a curvilinear function and an interaction was obtained because competition did not increase performance to the same extent at all levels of difficulty and was less effective for the more difficult problems.

Fang (1966) failed to obtain the expected interaction in a rather complex experiment. He used concept formation tasks with three levels of complexity, three levels of incentive ($1.00 in money, certificate of merit or no incentive) and two other variables. No statistical difference was found between incentive groups. Both this failure to obtain a difference between incentives and the lack of the anticipated interaction were explained in terms of insufficient length of the experimental session and an intrinsic interest in the task itself.

A series of studies by Spence, Taylor, and their associates (Spence, 1958) with eyeblink conditioning provided further evidence of a curvilinear relationship between motivation and performance. Level of anxiety, measured by the Manifest Anxiety Scale, and air puff intensity were the independent variables. A control group, comprised of randomly selected subjects to randomize anxiety level, was used for comparison. The positive, negatively accelerated curve partially supported the Yerkes-Dodson Law, but the lack of an interaction failed to support the primary part of the Law.

Generalization of predictions from the Yerkes-Dodson Law to human behavior is indicated by several studies, although a sufficient number of levels of the motivation variable have not usually been used with humans. This weakness in confirming studies, as well as the existance of studies failing to support the Law, indicate that its application to human learning needs extensive further study.

Summary of Research with the Yerkes-Dodson Law

A summary of the studies cited above was provided in Table 1. The most clearcut and consistent support for the Yerkes-Dodson Law has come from aversive motivation of small animals. The one study using alimentary motivation failed to support the Law. Studies with human subjects by Eysenck and his associates have provided some evidence of the curvilinear relationship between motivation and performance predicted by the Law and tentative evidence of the predicted interaction of motivation with task difficulty. Eysenck (1964) applied the Law as an explanation of the findings from his laboratory. An incompletely reported study by Wickens (1942) also indicated support for the Law, but a study by Fang (1966) failed to find support. Thus, although results are mixed, there seems to be enough evidence to indicate that the Yerkes-Dodson Law should be considered in any study of the interaction between motivation and task difficulty.

Spence-Taylor Drive Theory

Anxiety has been a basis for a large proportion of research into the motivation by task difficulty interaction. Generally the Test Anxiety Questionnaire (TAQ) or Manifest Anxiety Scale (MAS) have provided the basis for partitioning the subjects. Several theories have been developed which generally predict that on easy tasks, subjects with high anxiety will perform better than subjects with low anxiety, but the opposite will be true for difficult or complex tasks (e.g., Taylor, 1956; Child, 1954; Spence, 1958). The dominant theory that predicts this interaction has been that of Spence and Taylor. Table 2 presents a brief description of several studies bearing on this theory. The Spence-Taylor Drive Theory essentially was based on Hull's (1952) multiplicative function of stimulus intensity and drive. It was theorized that drive is a function of the strength of the emotional response made by an organism to noxious stimulation. The fundamental theory viewed response frequency as a positive monotonic function of excitatory potential (Spence, 1958). Anxiety, as measured by the Manifest Anxiety Scale, was defined as a drive for the purposes of evaluating the theory.

For simple classical conditioning studies the theory worked very well. A number of confirming studies have been conducted; performance by subjects high in measured anxiety is typically superior to those with low measured anxiety (cf. Spence, 1958).

The Spence-Taylor theory was applied to complex learning situations through the effect of drive on competing response hierarchies. Overt performance on any complex task was a consequence of complex interactions and competitions among correct and incorrect response tendencies (Brown, 1961).

Spence (1958) summarized the predictions of the theory for complex behavior: 1) if the habit strength of the correct response is greater than that of competing incorrect responses, increased drive level will produce a higher initial proportion of correct responses, shorter learning time, and fewer total errors. 2) if the habit strength of an incorrect response is greater than that of the correct

Table 2

Studies of Motivation and Task Difficulty Related to the Spence-Taylor Theory

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Spence (1956) Paired assoc. learning (several studies summarized)	Anxiety level (MAS) List difficulty	Percent correct responses	On easy lists, hi-A performed better than low-A; reverse on difficult lists (apparent disordinal interaction)	Support Spence-Taylor
Palermo, Castaneda and McCandless (1956) Difficult assoc. learning	Anxiety level (Children's MAS)	Mean errors	Hi-A group made consistently more errors than the low-A group	Partial support for Spence-Taylor
Castaneda, Palermo and McCandless Association learning	Association difficulty Anxiety level (Children's MAS)	Mean correct responses	Predicted disordinal (crossover) interaction. Main effects not significant	Supports Spence-Taylor
Standish and Champion (1960) P-A learning	Anxiety level (MAS) List difficulty	Reciprocal latency of response	Disordinal interaction	Supports Spence-Taylor

ERIC

Table 2 (Continued)

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Willett (1964b) P-A learning	Drive level (Eysenck) List difficulty	Response latency	No interaction; Hi-D had consistently lower latencies than low-D.	Fails to support Spence-Taylor
O'Neil, Spielberger and Hansen (1969) Spielberger (1970) Math problems	Anxiety level Problem difficulty	Mean errors Mean systolic blood pressure Mean anxiety state	Disordinal interaction; Anxiety data supported by data obtained during study	Supports Spence-Taylor
Hall (1970) Programmed instruction	Induced anxiety level Program difficulty	Posttest score Anxiety state score Program error rate	No interaction	Fails to support Spence-Taylor
DeBonis (1967) Concept association	Motivating instructions Association difficulty	Errors Response latency	Errors - nonsignificant disordinal interaction Latency - Hi-D had consistently shorter latency than Low-D	Partial support for Spence-Taylor

Table 2 (Continued)

Investigator/Task	Independent Variables	Dependent Veriables	Results	Comments
Feldman (1964b) Digit cancelling	Difficulty level Drive level (Eysenck)	Mean correct responses Mean errors	No interaction	Due to nature of the task, the results may be considered in partial support of the Spence-Taylor Theory

response, an increase in drive level will result in poorer performance initially, but eventually the habit strength of the correct response will dominate and, from that point, the proportion of correct responses should be higher for the high drive group. It should be noted that the theory states that "eventually" the habit strength of the correct response will dominate; in most studies, this point is not reached. In effect, the theory predicts a crossover (disordinal) interaction between drive level and task difficulty (it is assumed that for tasks of high difficulty the habit strength of incorrect responses is typically higher than that of the correct response). For easy tasks an increase in drive will yield an increase in performance; the reverse will occur for difficult tasks.

A recent reinterpretation of the findings from studies of test anxiety has been made by Wine (1971). Wine suggested that the critical variable is the division of attention paid to self-relevant and task relevant variables. The adverse effects of high test anxiety are due to a division of attention between the variables whereas subjects with low test anxiety focus their attention more fully on the task. The predictions from this proposal seem to be the same as those from the Spence-Taylor theory.

The Spence-Taylor Drive Theory has led to a moderate number of studies of the interaction between motivation and task difficulty. Spence (1956) reported several studies that support the theory. Paired associate learning tasks of different levels of difficulty (based on degree of response competition) were crossed with two levels of anxiety, measured by the Manifest Anxiety Scale (MAS). The typical finding was that for easy lists subjects with high measured anxiety had a higher proportion of correct responses than those with low measured anxiety. These results were reversed for difficult tasks.

Palermo, Castaneda, and McCandless (1956) found that subjects with a high level of anxiety measured by a children's form of the MAS made consistently more errors than those with a low level on a difficult association-learning task. Castaneda, Palermo, and McCandless (1956) crossed association learning tasks of two levels of difficulty with anxiety (children's form of the MAS) and found results similar to those reported by Spence (1956).

Studies of the interaction between motivation and paired associate learning list difficulty by Standish and Champion (1960) and Willett (1964b) yielded conflicting results. Both used response latency as a dependent measure. Standish and Champion partitioned their drive groups on the basis of MAS results and obtained the interaction predicted by the Spence-Taylor Theory. Willett used the operationally different groups described above by Eysenck (1964): his high drive group yielded lower response latencies than the low drive group at both levels of difficulty.

Atkinson (1964) has pointed out that there is little evidence that subjects with high anxiety perform better on easy tasks than those with low measured anxiety, although the results seem to be mixed. Feldman (1964a) has criticized the MAS (and paper and pencil measures of anxiety in general) as intrinsically unsatisfactory for measuring drive level. He pointed out that correlations between the MAS and clinically anxious and non-anxious patients were unsatisfactory and agreed with Lawrence (1958) that "...direct manipulation of drive seems more appropriate than the indirect selection of subjects by paper and pencil test" (p. 166). Travers (1963) found no relationship between the MAS and empirically determined anxiety.

Some researchers have attempted to exert better control over anxiety. For example, O'Neil, Spielberger, and Hansen (1969) buttressed their argument for a difference in anxiety, measured by the State-Trait Anxiety Inventory, by obtaining systolic blood pressure measurements that confirmed a difference in excitement. Using the number of errors on mathematics problems as a dependent variable, they obtained the predicted disordinal interaction at an acceptable level of significance. Spielberger (1970) replicated the study with the same results. Hall (1970), on the other hand, attempted to create stress (anxiety) by instructions. The State-Trait Anxiety Inventory was administered during the experiment to obtain information concerning the level of induced anxiety. Programmed instructional materials, one with a normal error rate (.05) and one with a very high error rate (.54), were used. The author thought that the results of this study were not significant primarily because the motivational manipulations failed to produce an adequate difference in anxiety between the groups.

DeBonis (1967) used two levels of instructions and two levels of difficulty of a concept association task. She verified the initial levels of homogeneity of anxiety in her four groups by testing with the Manifest Anxiety Scale for Children, then she induced new levels of anxiety with the instructions. Although her design was weak, the predicted curves for a disordinal difficulty-by-motivation interaction were obtained with number of errors as the dependent variable. However, the F-ratio failed to reach significance. Her hypothesis was based on the predictions of Taylor's theory and that of Yerkes and Dodson.

The results of Eysenck (1964) and his associates, described above, may be applicable to the Spence-Taylor Theory (see Table 1). The set of three experiments with tasks of different difficulty resulted in superiority of the high drive group over the low drive group on the easy task and the reverse with a complex task. Although the tasks and the dependent variables differed, the absolute magnitude of difference between groups performing the easy and complex task was greater than the magnitude of difference obtained from groups working on a task of inordinate difficulty. These studies may be viewed as providing support for the Spence-Taylor theory. However, another of Eysenck's co-workers failed to obtain supporting data (Feldman, 1964b). He used the two operationally defined drive groups described by Eysenck above (applicants and employees) with a task consisting of cancelling digits (two levels of difficulty). He found that the mean number of correct responses favored the high drive Ss at both levels of difficulty, but both drive groups made more errors at both levels. However, it might be argued that the task did not give rise to competing response hierarchies.

<u>Summary of Spence Taylor Drive Theory Research</u>. The Spence Taylor Drive Theory has received reasonably consistent support from studies using anxiety measured by the Manifest Anxiety Scale, as the drive variable. The results of studies using other drive variables have been mixed. With the questions of the

validity of the MAS that have been raised it appears that the validity of the theory has not been firmly established. However, the results of the recent, well-conducted experiments by O'Neil, Spielberger and Hansen (1969) and Spielberger (1970) indicate that the theory is still deserving of recognition.

Although the Yerkes-Dodson Law and Spence-Taylor Theory seem quite different and stem from different approaches to learning, there are some essential similarities. The motivations fundamental to both are essentially negative in that there is an element of desire to avoid or escape. Both approaches predict an interaction between motivation and task difficulty, although the Spence-Taylor theory specifically predicts a disordinal interaction and the Yerkes-Dodson Law does not, although it might occur. The differences between the approaches are important, however. Spence-Taylor is concerned with complex learning and is based on a theoretical foundation. Yerkes-Dodson is empirical in nature and has been primarily applied to simple conditioning processes. The curvilinear relationship between motivation and performance that is fundamental to Yerkes-Dodson has been indicated by the results from studies of complex learning, but data on the nature of the motivation-task difficulty interaction is still lacking due to the limited number of levels of motivation that are typically used.

Atkinson's Theory

Atkinson (1964) has developed a multiplicative theory of achievement motivation that is based on a general motive to achieve success (M_s), the expected probability of achieving success (P_s), and the incentive value of success (I_s). The tendency to achieve success (T_s) was defined as a multiplicative function of these components: $T_s = M_s \times P_s \times I_s$. To reflect the incentive value of difficult tasks, he defined $I_s = 1 - P_s$. He also took account of the motive to avoid failure by defining the incentive to avoid failure (I_f) as the negation of P_s: $I_f = -P_s$. It was tentatively assumed that the expectancy of failure (P_f) would be weak when the expectancy of success (P_s) is high: $P_s + P_f = 1$. The tendency to avoid failure (T_{-f}) was seen as inhibitory, which differed from the approach of Taylor and other theorists. T_{-f} was jointly determined by M_{af} (motive to avoid failure), I_f, and P_f, so that $T_{-f} = M_{af} \times I_f \times P_f$. Thus, the resultant motivational tendency was the sum of the tendency to achieve success and the tendency to avoid failure: $T_r = T_s + T_{-f}$. To this, Atkinson also added an extrinsic motivational variable, T_{ext}, so that the total strength of tendency $T_t = T_r + T_{ext}$.

Based on his theory, Atkinson (1964) predicted a complex curvilinear relationship between task difficulty (defined in terms of perceived probability of success), motivation (the "resulting tendency") and performance. Subjects with high n Achievement and low n Avoid Failure would produce performance data represented by an inverted U-shaped curve with a peak at $P_s = .5$; those with low n Achievement and high n Avoid Failure would have a U-shaped curve with a minimum at $P_s = .5$. These are shown in Figure 2.

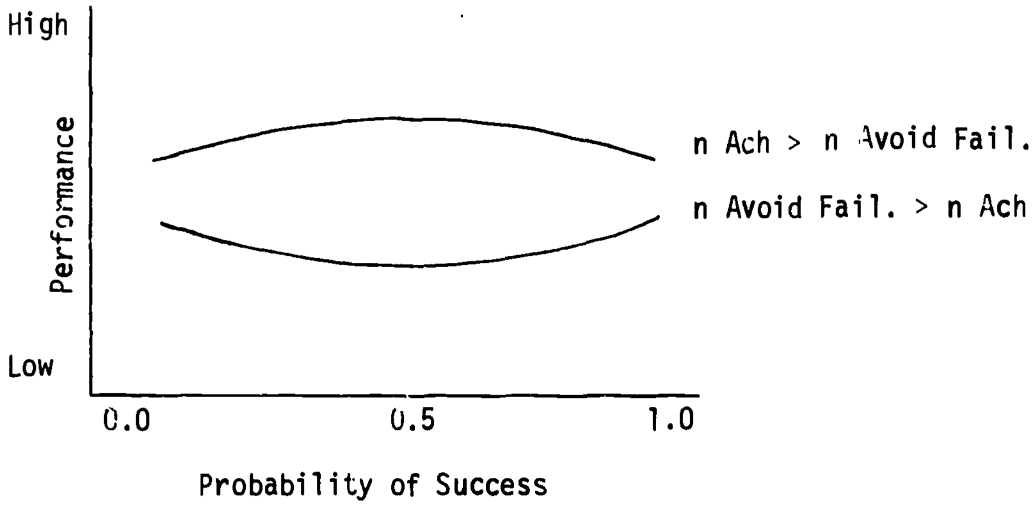


Fig. 2. Graphic representation of Atkinson's Theory. (Based on Atkinson, 1964)

Several studies have been found that are relevant to Atkinson's theory. These are listed in Table 3. Several instruments have been used for the assessment of achievement and failure avoidance motivation. Prominent among these are the projective Thematic Apperception Test (TAT) and objective measures such as the Test Anxiety Scale (TAS) and Mandler-Sarason Test Anxiety Questionnaire (TAQ).

Atkinson and Litwin (1960) reported the results of a ring toss game with different throwing distances in which the high n Achievement-low Anxiety (which was equated with n Avoid Failure) subjects had the highest peaks, those with reverse motivational characteristics had the lowest. It should be noted that curves for both motivational relationships were roughly normal, or Gaussian, in form. Karabenick and Youssef (1968) obtained the predicted curves with paired associate tasks of equal difficulty but with different established subjective probabilities of success (easy, intermediate, and difficult). A significant motivational group by difficulty interaction was obtained. These two studies provide strong support for the theory.

Table 3

Studies of Motivation and Task Difficulty
Related to Atkinson's Theory

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Atkinson and Litwin (1960) Ring toss	n Achievement Anxiety Throwing distance	Percentage of shots	Apparent interaction; curvilinear relationship between difficulty and performance at both motivational levels	Held to support Atkinson's Theory
Karabenick and Youssef (1968) P-A learning	Subjective prob. of success n Achievement n Avoid Failure	Mean number correct	Predicted curvilinear relationship obtained	Supports Atkinson
Kight and Sassenrath (1966) Programmed materials	n Achievement Anxiety	Completion time Error rate Retention test scores	Hi n Ach-Hi A superior to Lo n Ach-Lo A on dependent variables	Fails to support Atkinson
Shrable and Sassenrath (1970) Programmed materials	Replication of Kight and Sassenrath		Complex but negative relationship between n Ach and performance	Appears to support Atkinson

Table 3 (Continued)

Investigator/Task	Independent Variables	Dependent Variables	Results	Comments
Entin (1968) Mathematics problems	Problem difficulty n Achievement n Avoid Failure Induced motivation level	Mean performance	Parallel curves for both levels of difficulty	No apparent support for Atkinson; partial support for Yerkes-Dodson
Feather (1961, 1963) Figure tracing problems	n Achievement Anxiety Subjective prob. of success	Number of trials	Significant interaction between motivation and subjective prob.	Held to support Atkinson's Theory
Maehr & Videbeck (1968) Cognate guessing	Risk-taking tendency Probability of success	Number of trials (persistence)	Inverted U-shaped curves for both levels of risk-taking	Held to support Atkinson's Theory if task was assumed achievement-oriented
DeCharms and Carpenter (1969) Spelling and arithmetic problems	n Achievement Risk-taking	Percentage of problems chosen at each risk level	Apparent interaction; curvilinear relationship predicted by Atkinson for both tasks	Supports Atkinson's Theory

ERIC

N. T. Feather, a student of Atkinson's, conducted a series of studies to study the application of the Atkinson theory to task persistence (Feather, 1961, 1963). His tasks consisted of figure tracing problems. Number of trials (attempts) to solve an insoluble problem before turning to an alternative was used as the dependent measure. High and low subjective probabilities of success were given subjects for each figure. He found that achievement-oriented subjects persisted longer at the insoluble task if it was represented as moderately easy ($P_s = .70$); failure avoidance-oriented subjects persisted longer if the task was represented as difficult ($P_s = .05$). This was held to support Feather's hypotheses and Atkinson's theory.

Maehr and Videbeck (1968), in a study of the effects of risk-taking tendency on preference for reinforcement levels, found that subjects were, overall, most persistent at moderate (50%) reinforcement-success schedules and least persistent at the extreme (15% and 85%) schedules. Because the task was of an achievement nature and subjects were presumed to be high in success motivation, the results were held to support Atkinson's theory.

Additional support for Atkinson's theory was obtained by deCharms and Carpenter (1968) with a different kind of study. The interaction of achievement motivation with risk-taking behavior was assessed with arithmetic and spelling problems. Subjects were fifth and seventh grade Negro children. Problems were developed to form six levels of difficulty for each kind of task, ten problems constituted each level. Subjects were pretested with a standard set of sixty problems in each set. The task consisted of solving problems and obtaining as many "points" as possible. The number of points possible for a given problem was based upon the difficulty level of the problem and the performance on the pretest. Each subject selected ten problems from each kind of task. A risk-taking element was inherent in each choice: the more difficult the problem selected, the greater the number of points that could be earned if successful. The results showed significant differences between the high and low n Achievement groups for both tasks. Curves plotted for the n Achievement by risk interaction were very close to those predicted by Atkinson's theory. In fact, the predicted U-shaped curve was found for the low n Achievement group in Spelling.

Conflicting results using programmed instructional materials have been reported by Kight and Sassenrath (1966) and Shrable and Sassenrath (1970). Using "easy" programmed instructional materials, Kight and Sassenrath (1966) found that high n Achievement - high Anxiety subjects worked faster and made fewer errors than low n Achievement - low Anxiety subjects, and high n Achievement subjects had higher retention scores. These differences were statistically significant. According to Shrable and Sassenrath (1970), Atkinson's model would predict either an approximately equivalent performance or a negative relationship between achievement motivation and performance level on this task in which the probability of success was .95 or better. Shrable and Sassenrath (1970) repeated the experiment with higher level college students and obtained the negative relationship. Subjects with high n Achievement - low Anxiety scores made the greatest number of errors on the program and tended to score lower on the retention test than the low achievement group.

ERIC

Entin (1968) used projective tests to assess n Achievement and n Avoid Failure. The tasks consisted of two-digit addition problems and Wendt three step problems. A difference in motivation was induced by specifying that the results would be made public or would remain confidential. He found a curvilinear relationship between motivation and performance for both levels of difficulty, with the groups assumed to be intermediate in strength of motivation having the highest mean performance at both levels. Although this study fails to support Atkinson's formulation, it might be seen as providing partial support for the Yerkes-Dodson Law.

A recent review of Atkinson's theory by Maehr and Sjogren (1971) cited several of the above and other studies in support of the validity of the theory and its employment in academic motivation situations and persistence behavior. These authors observed that the obtained evidence indicated inverted U-shape functions for both n Ach > n Av Fail and n Av Fail > n Ach, but the mode of the former was greater than that of the latter, indicating that both exhibit a preference for moderate difficulty levels although achievement-oriented subjects prefer it more than failure-threatened subjects. Maehr and Sjogren pointed out that it remains to be demonstrated that failure-threatened subjects exhibit maximum avoidance in the moderate difficulty range, as predicted by the theory. However, the obtained results may be due to some unspecified motivation acting as an additive factor.

The criticisms that were levelled against the anxiety studies should also be valid for the above studies. The primary means of assessing n Achievement and n Avoid Failure are the use of projective-type tests (e.g., the TAT) and tests of anxiety (e.g., MAS or TAQ).

Birney, Burdick, and Teevan (1969) point out that the use of the TAQ as a measure of fear-of-failure is wholly dependent upon the theory that n Avoid Failure is defined by low n Achievement and high scores on the TAQ. Maehr and Sjogren (1971) criticized the instruments used to measure achievement motivation and motivation to avoid failure and pointed to the need for the development of better instruments. Clearly, direct manipulation of the motivational variables may be most desirable.

Summary of Atkinson's Theory of Achievement Motivation

Atkinson's Theory is attractive because of its attempt to quantify the motivation-task difficulty relationship. The theory permits predictions for a continuum of task difficulties, as the Yerkes-Dodson Law permitted predictions for a continuum of drive levels. However, Atkinson's theory is limited to only two motivational variables; he does not appear to have attempted to generalize directly to other variables. Indeed, the motives of achievement and failure avoidance do not appear to have been manipulated; they have just been measured.

The second independent variable of Atkinson's Theory is task difficulty (probability of success). Studies that have supported the theory have used a sufficient number of levels of difficulty to allow the predicted results to appear. Studies that have not supported the theory or on which conflicting evidence was obtained used few levels of difficulty. Because this study employed several difficulty levels, resulting data should have a bearing on the theory.

Atkinson's Theory, like the other two approaches described above, leads to a prediction of an interaction between motivational variables and task difficulty. However, the nature of the interaction is different. Atkinson does not predict a disordinal interaction, although it would appear from the nature of his predicted curves (Figure 2) that such an interaction might be obtained under certain circumstances if two levels of difficulty are used.

The above review of two theories and a law indicates that a great deal of research still needs to be done to clarify the relationship between motivation and task difficulty. The two theories reviewed are tied to specific motivational variables. The Yerkes-Dodson Law, although presumably generalizable to all motivational variables, has not been adequately tested with human subjects.

Hunt's Hypothesis

Hunt (1961, 1965, 1971) has hypothesized that the differences in task difficulty or complexity may, in themselves, create differences in intrinsic motivation. Thus, a student may feel challenged more by difficult problems than by easy problems and work longer, faster, or more accurately on them. Little research has been found that bears on this theory, but if the hypothesized effect exists, it should have a major impact on the results of studies related to Atkinson's Theory. Although his earlier formulations have been primarily applicable to very young children, Hunt's recent statement (1971) indicated applicability to children of all school ages. He pointed out that a critical element in finding the appropriate level of complexity is the child's freedom to select his model, to take it or leave it. The existence of such a possibility clearly calls for the use of control groups that receive no external motivation as an added dimension of the interaction research, and may have accounted for the curvilinearity found in the isolated (control) group by

Wickens (1942). However, in many studies this control is lacking and there is no evidence concerning any possible interaction between the hypothesized intrinsic and the imposed extrinsic motivation. This seems to be a rich area for research, but little has been found.

Summary of Interaction Studies

Based on the results of Yerkes-Dodson (1908) and others, and upon Hebb's (1955) formulation, it seems logical to postulate a curvilinear relationship between motivation and performance. Following Yerkes and Dodson, it seems reasonable to expect that the optimum motivation may differ with the difficulty of the task. Furthermore, Atkinson and Hunt postulate a curvilinear relationship between task difficulty and performance. When one is dealing with a curvilinear phenomenon it is necessary to design an experiment that will reflect the .rve that is expected. Those studies that found nonsignificant negative results might be explained by a failure to insure that the design adequately took account of a possible curvilinearity. This would be especially true of simple 2 x 2 factorial interaction designs. One weakness of the Spence-Taylor approach is the inherent lack of consideration given to the possibility of curvilinear relationships between their variables and their dependent measures.

The ideal test of the approaches described above would involve several levels each of the motivation and task difficulty variables. Several kinds of each variable should be tested. However, this would require a very large number of subjects and several years to accomplish. The present study was designed to provide a start for such an endeavor by evaluating two educationally significant variables, incentive (the motivational variable) and arithmetic problem difficulty (the task variable). The rationale for the selection of incentive and research concerned with incentives and with task difficulty will be described in the next section.

Evaluation of Studies
Of Motivation and
Task Difficulty:
Studies of Task Difficulty

Some studies of task difficulty have been conducted, but generally without consideration of individual difference variables. For example, a large-scale study reported by Blake (1964) evaluated performance on a variety of tasks with several dimensions of complexity. In general, he found that performance was inversely related to task difficulty, that IQ and sex influenced learning and that method of task presentation influenced learning in only a few situations. These seem to be typical findings, although Fang (1966) points out that differences in experimental findings in this area have often been traced to differences in the characteristics of the tasks used. There seems to be a serious lack of useful operational definitions of complexity and difficulty that are generalizable across tasks. There is a need for the development of an operational definition of task difficulty that can be used to specify levels, or degrees, of difficulty that are individualized to each subject rather than based upon normative standardization.

The Need to Study Rewards or Incentives as Motivational Variables

A large proportion of studies dealing with motivation and task difficulty have used anxiety or related constructs as the motivational variable. These motivational variables have some relevance to the school learning situation. Certainly, if Skinner (1965) is to be believed, aversive stimuli that increase anxiety are the dominant means of motivating students. Even if Skinner is wrong, it cannot be denied that fear of failure is a powerful motive for at least some students to perform well in their academic work, and during a period of intense academic competition perhaps it is the dominant stimulant for a majority of students.

An alternative to the threat of failure might be the reward of success. It can be argued that high academic grades are a reward if they are achieved, but they are found on the same continuum as a failing grade; the threat of failure always exists. As was pointed out earlier, praise is frequently used but its effect on learning is not really known. However, the success of contingency management systems in a variety of settings shows that tangible reinforcement and reward techniques can be highly effective in shaping human behavior and in encouraging desirable study behaviors.

Studies of the Effects of Incentives

A small number of studies have been found that incorporate incentives into the experimental design. "Incentive" refers to the expectation that a particular reward will be obtained after behaving in a certain way (Cartwright and Cartwright, 1970), usually upon the completion of a series of responses or at the end of the experimental period. These studies have generally failed to yield clearcut results concerning the efficacy of incentive. Goodyear (1970) promised a reward of three or ten points added to a semester grade for good performance on an immediate recall and reflective listening test. Other groups were told they would have three or ten points deducted if their performance was below a specified criterion level. A control group was told their performance would have no impact upon their grade. He found no evidence that reward (as he operationally defined it) produced any significant effect.

Goyen and Lyle (1971a, 1971b) compared money and knowledge of results with no feedbacks or incentives for two different tasks with normal and retarded readers. No significant differences were found for tachistoscopic recognition of shapes (Goyen and Lyle, 1971a), but significant differences favoring the incentive groups for both reading ability groups were found when the task was one of learning geometric figure associations (Goyen and Lyle, 1971b). The authors proposed that the difference in effect was due to the difference in tasks.

A set of twelve incentive objects was scaled and tested by Haaf, Feldstein and Witryol (1970). They found three clusters of preference for objects that were common among sex and grade level groups tested. Haaf (1971) used this scale with a form discrimination problem. Specific rewards were correlated with position (left-right) if the correct stimulus was in that position. Although there were no significant differences between reward groups, he did find indications of differential effectiveness of the rewards. However, the relative effectiveness did not agree with the results of the previous scaling.

Noffsinger (undated), with a visual tracking task, found that normal subjects performed better with no reward than with the incentives of money or verbal reward. This seems to contradict the findings of the studies by Stennett (1957) cited above.

Clifford (1971) evaluated the effects of a reward with homogeneous and heterogeneous group competition. A substitution task was used. She found the homogeneous competition with reward group to be superior, with reward making little difference in performance for the heterogeneous group competition subjects. However, she also found a clear ability effect: high ability subjects showed a greater resistance to reward than low ability subjects. She pointed out that the results "raise the question of whether success and/or recognition of success may have a relative value dependent upon subgroup status." (p. 14).

From the above study, it appears that the efficacy of an incentive may depend upon the nature of the task, the nature of the reward, the ability of the subjects, and the value of success to the subject.

A review by Fang (1966) found five studies in which incentives facilitated performance on certain types of tasks and three in which incentives had no effect. He also reviewed studies which attempted to compare the efficacy of different types or levels of reward: a difference was found by one investigator, no difference by three.

The contingency management studies, which seem to be gaining favor for educational and non-educational applications, utilize a variety of incentives that can be obtained by accumulating tokens, points, or other reinforcements for specific behavior (e.g., Lackner, 1970; Wrobel and Resnick, 1970; and Breiling, Shipman, Milligan and Pepin, 1970). The student may select any reward from a menu as long as he has accumulated the required number of points or tokens. However, no studies of contingency management programs have been found that attempt to evaluate individual preferences for different kinds of incentives.

Reward Preference Inventory

Perhaps the best effort to quantify reward preferences is that of Dunn-Rankin (e.g., Dunn-Rankin and Shimizu, 1968) with the development of the Reward Preference Inventory (RPI). This instrument, which is described in detail below, is an attempt to obtain the reward preferences of children by an indirect, paired comparison technique and to obtain preferences across a wider range of categories than has been attempted previously. The categories consist of adult approval, competition, peer approval, independence and consumable. The RPI has been carefully developed and validated by administration in several schools at widely separated geographic locations. Cartwright (1968, 1970) used the RPI to study the result of using incentives of high and and low reward value on performance of a paired-associate learning task and a perseverance task. Although significant differences between the incentive groups were not found, the trend was in the expected direction (high incentive superior to low incentive) for the number of P-A trials required for learning and for time spent on the perseverance task. However, it is interesting to note that the number of errors on the P-A task was lower, and the number of correct responses on the perseverance task was higher, for the low incentive groups. These findings are similar to those of Feldman (1964b), Eysenck (1964), and Palermo, Castaneda and McCandless (1956), and indicate that the motivational variable used in this study may be equated with other motivational variables. Additional evidence of a difference between the incentives was found in that seven subjects in the low incentive condition spontaneously rejected the promised reward.

Confounding effects due to extraneous variables were found by Cartwright (1968, 1970) which might explain the failure to obtain significant results. Significant correlations were obtained between four variables (reading, arithmetic, achievement, and IQ) and the standard score on the perseverance task. It was observed that differences in achievement were sufficient to mask any effects due to the experimental variables. However, the evidence seems to indicate that there is a motivational difference between the high and low incentive groups. Cartwright has observed that, ". . . the Reward Preference Inventory is an experimental instrument; research directed toward establishing predictive validity across a wide range of learning situations and for varied groups of individuals in needed," (1970, p. 157).

In order to evaluate the usefulness of the RPI in a computer assisted instruction (CAI) environment, a series of pilot studies (unpublished) have been conducted under the direction of

G. P. Cartwright that replicate in a CAI setting the study by C. Cartwright (1970). The independent variable was the promise of the most preferred or least preferred reward; the dependent variables were time spent on a repetitive task of finding 5's and the number of 5's found.

The pilot study with about thirty children, aged eight to ten, has led to the following conclusions concerning the use of the RPI in a CAI environment: 1) it is difficult to operationally provide effective peer approval or independence rewards; 2) children tend to forget the reward they were promised; and 3) the operation of a CAI terminal is, in itself, a strong incentive.

It was found that only a very small number of children selected the peer approval or independence rewards. An analysis of the rewards actually offered, a certificate to be shown to a best friend (for peer approval) or free play time (which all subjects received anyway, for independence) were not perceived as being equal in value to the other offered incentives (candy bar, certificate of merit, name on public bulletin board). No effective way of increasing the value of the least favored categories has been found and it was not thought desirable to decrease the value of the others.

In order to strengthen the effect of the RPI, based on the pilot studies described above, the following modifications were made in an additional pilot study: 1) the Peer Approval and Independence categories were eliminated; 2) additional pairings of examples of the remaining three categories were included for a new total of 24 pairings; and 3) intermittent statements of encouragement were added to the task. These encouraging statements have special reference to the particular reward assigned to the subject and are designed to remind him of the reward for which he is working.

Preliminary results using this modified method indicate that it results in a larger difference in the dependent variables between the most and least preferred reward groups than did the original procedure.

Rationale and Anticipated Value

The Yerkes-Dodson Law, Spence-Taylor Theory, and Atkinson Theory have used, and been based upon, different motivational variables. Although some studies have been found that may be applied to two of the approaches, none have attempted to determine whether any of the theories may be generalized to include motivational variables other than those specified by the theory. In this study, the "strong inference" technique suggested by Platt (1964) was invoked by selecting a motivational variable that was implied by none of the theories but which might permit generalization to all the theories. Although, as Platt pointed out, such research is high risk, it was felt that it was necessary to start searching for generalizations within the motivation by task difficulty paradigm.

Studies related to the theories have typically used levels of difficulty based upon normative data or upon specified a priori probabilities of success (although the actual problems frequently bore little similarity to the specified probabilities). However, normatively based differences in a task do not take account of the individual differences of subjects; what may be difficult for one subject may be relatively easy for another. A major question that might be asked of these studies is whether control of the difficulty level for each subject would yield the same results. One intent of the present research was to evaluate such an individualized difficulty level within the motivation-by-task difficulty interaction paradigm.

Incentive was selected as a motivational variable for three reasons. First, it met the requirement of not being identified with any of the approaches being evaluated. Second, it is a manipulatable variable of educational significance in that it might be applied in educational environments to stimulate improved performance of students. Third, an extensively validated instrument exists, the Reward Preference Inventory, that would provide an indirect assessment of individual preferences for rewards and permit the assignment of subjects to high and low preference conditions.

From both a theoretical and a practical standpoint, it seems appropriate and worthwhile to investigate the interaction of incentive motivational variables with task difficulty using computer assisted instruction equipment. The computer can be an effective instrument for maintaining specific levels of difficulty adapted to each student, for collecting information concerning reward preferences, for assigning students to reward conditions, and for providing information to the experimenter concerning the reward status of each subject to minimize human influence. The feasibility of using a computer for such a project has been demonstrated by O'Neil, Spielberger and Hansen (1969) at Florida State University and by pilot studies with the Reward Preference Inventory conducted at Penn State.

Data bearing on several theories will be obtained by a study of the type proposed. If sufficient difference between strengths of incentives is obtained, the theories can be pitted against each other to determine the validity of each. From the proposed wide range of levels of difficulty, a test of Hunt's (1961, 1965) hypothesis of an optimum difficulty level for intrinsic motivation can be conducted. From the use of the RPI, further evidence of the validity of this instrument can be obtained. Furthermore, useful information should be provided to authors of CAI drill and practice programs for whom no real guidelines exist concerning the specification of problem difficulties.

ERIC

CHAPTER III

PROCEDURES

Definition of Independent Variables

Incentive level. Preference for types of incentives was determined with the modified Reward Preference Inventory (RPI) (Appendix B contains a complete description of the text). The results of the RPI provided a quantitative measure of preference for three categories of reward based on a series of paired comparisons between samples of adult approval, competition, and comsumatory response and permitted the division of subjects into two groups. In one group, each subject received his most preferred reward; in the other, the least preferred reward was received. The Most Preferred Reward (MPR) was defined as the category for which the subject indicated a preference the largest number of times. The Least Preferred Reward (LPR) was the category selected the fewest number of times. A control group received the RPI but was not told that a reward would be received. Subjects were randomly assigned to the three reward groups. Those receiving a reward were told what they would receive when they finished, but were not specifically told that the reward was their highest or lowest preference.

The specific incentives used were the following:

a. Adult Approval-Report card with an A entered as the grade.
b. Competition-Name of subject posted on a chart indicating excellence prominently displayed in the terminal room.
c. Consumatory-M&M candies.

Replicas of the report card and chart are in Appendix C.

Task difficulty. A task variable was sought that would permit several levels of difficulty to be operationally defined. Arithmetic problems were selected because it was possible to meet this criterion by proper programming of the computer. There have been several precedents for using arithmetic tasks on computer assisted instruction equipment (e.g., Suppes, Jerman and Brian, 1968; and recent studies by Heimer and his students at the Pennsylvania State University[1]).

Difficulty level was operationally defined in terms of the percent of problems solved correctly. Four ranges of percent correct (hereafter referred to as PCR, for Percent Correct Range) were specified. The ranges selected were 98-80%, 72-54%, 46-28%, and 20-2%. The computer could adjust the kind of problem received to insure that subjects' performance stayed within the prescribed range. In this way, the difficulty of the task was maintained within prescribed limits on an individual basis, thus taking account of differences in ability. It should be noted that this procedure for handling task difficulty is quite different from that typically used for studies of the Spence-Taylor Theory and Yerkes-Dodson Law, but is somewhat analogous to that used for tests of Atkinson's Theory.

Dependent Variables

Sensitive dependent variables were required because it was expected that the differences in motivation produced by the differences in incentive value might not be of a large magnitude. Also, it was

---

[1]Personal communication.

expected that a large amount of variation in the results might be produced by utilizing a computer system for the presentation of stimuli. Previous studies with the RPI using a perseverence task used working time, standardized number of solutions, number of errors, and solution rate as dependent variables. The studies used open-ended tasks which seemed to be quite sensitive to differences in motivation, and Feather (1961, 1963) found persistence at a task to be a good measure of the strength of competing behavioral tendencies. He used time spend on each task and number of trials as his dependent measures.

Selecting dependent variables for a study of this type is difficult because the relevant theories assume different dependent variables as well as different independent variables. The Yerkes-Dodson Law and Spence-Taylor theory assume measures of learning as dependent variables; Hunt's theory seems to imply both learning measures and persistence measures. Both persistence and performance are inherent in Atkinson's formulations. Clearly, individualized control of the level of difficulty does not permit direct evaluation of learning. For the Yerkes-Dodson Law this is relatively unimportant. Because this approach is empirical in nature it should be expected that the effect would manifest itself whether the dependent measure is number of trials to criterion or amount of time spent on the task, although it was developed to account for learning data. However, the change in the nature of the dependent measure is an important consideration for the Spence-Taylor theory.

The Spence-Taylor theory postulates that differences in motivation will result in differences in the way competing response hierarchies are handled, e.g., whether the correct answer becomes dominant easily or not. It is expected that for easy tasks in which the correct answer is dominant subjects with high motivation will persist longer at tasks with which they are having success than those with low motivation. However, for difficult tasks where incorrect responses dominate the response hierarchy, it is expected that subjects with high motivation will experience severe frustration earlier than those with low motivation. Severe frustration should result in termination of the task; thus, subjects with high motivation who encounter difficult problems should persist for a shorter time than those with low motivation. This leads to a prediction of results with a persistence variable that are similar to those from the learning variable assumed by the theory and provides a basis for using persistence as a dependent variable for evaluating the results from the study in terms of the Spence-Taylor theory. It appears, then, that persistence measures might provide the needed extrapolation of results to the evaluation of the interaction theories and can also permit comparison with prior studies conducted with incentive as an independent variable.

The selection of specific dependent variables for the present study was complicated somewhat by the fact that different amounts of time would be required to solve different kinds of problems. For example, it was clear that any subject could solve single digit addition problems in much less time than multi-digit multiplication problems.

The following two dependent measures were selected on the basis that they contained the least bias as measures of persistence and they provided comparability with prior studies by other investigators: 1) time spent working on the task (TTime) and 2) an adjusted number of problems attempted (ANP). TTime represented only the time spent on the task problems and was calculated from times given in the student records output.

Several measures of task performance were obtained, including number of correct answers, number of wrong answers, and number of overtimes. Subjects were also given the option of typing the letter "n" if they did not know how to do a problem. During subject trials, it was determined that several subjects used the "n" to try to obtain a different kind of problem. As a result, it was decided that the most appropriate measure of the number of problems really attempted would be the total number of problems attempted minus the number of "n" entries. This constituted the ANP variable selected as the second dependent measure.

Equipment

Subjects used terminals of an IBM Instructional System mounted in a mobile van. Each terminal consisted of an IBM 1510 keyboard and display screen with a photo-sensitive light pen and an IBM 1512 Image Projector. All materials were presented via these terminals. Materials were written in the Coursewriter II language.

Illustrations of a representative terminal and the interior of the van are provided in Figures 3 and 4. A detailed description of the computer based instructional system is given in Appendix D.

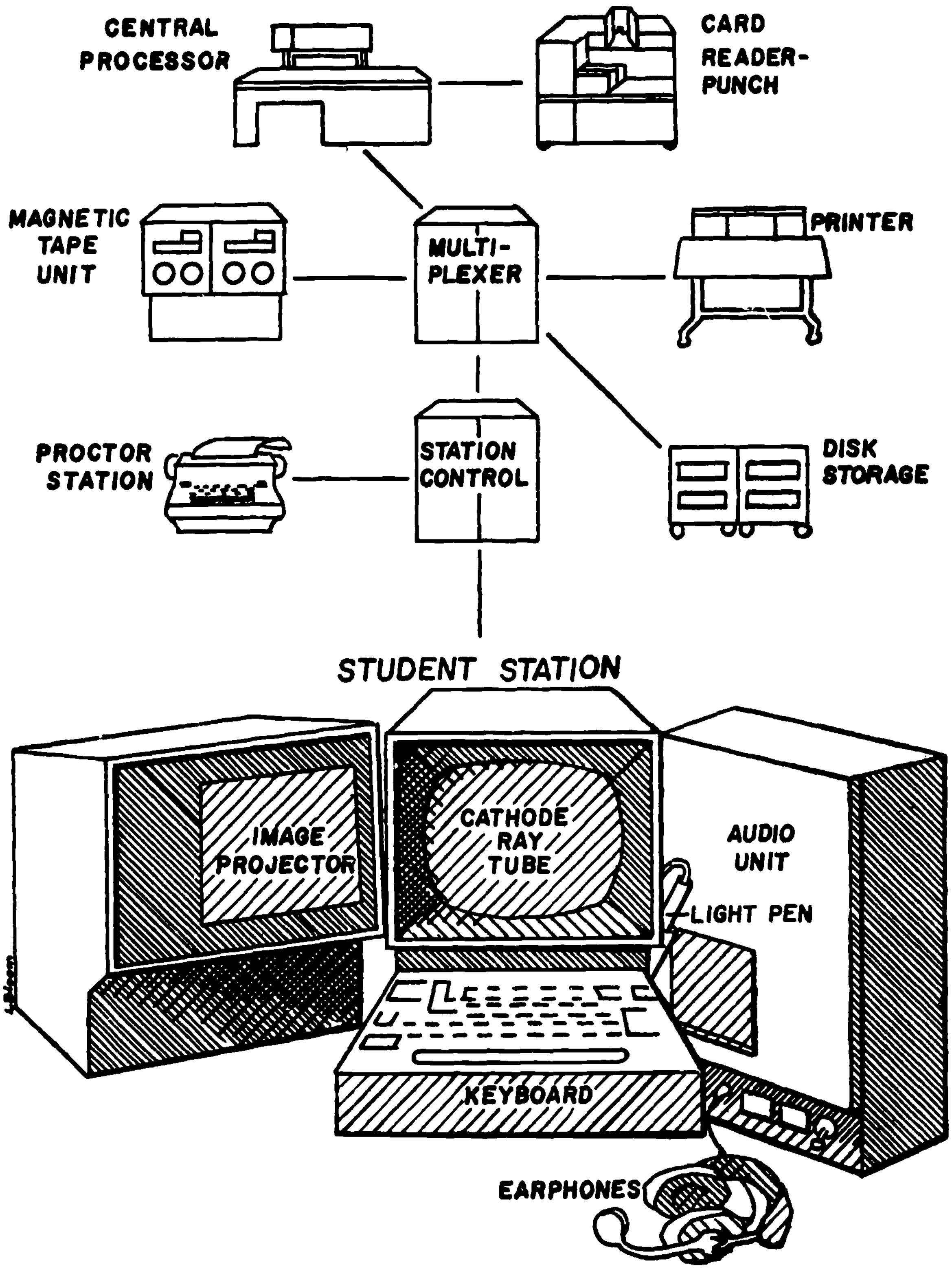


Fig. 3. Illustration of the Instructional Terminal and System Components of the IBM 1500 Instructional System.

Fig. 4. Interior of the Computer Assisted Instruction Mobile Van.

Subjects

Subjects were selected from fourth and fifth grade classes at the Athens Main Elementary School in Athens, Pennsylvania. Athens, a town of approximately six thousand persons, is located in a small complex of towns just south of the New York State border near Elmira, New York. The principle economic activities in the Athens district are agriculture and light industry.

Prior to initiating the study, teachers were familiarized with the intent and content of the experiment. Ten teachers in the fourth and fifth grades, representing about 300 students, agreed to permit their students to participate. Due to logistical difficulties, a critical time variable, a desire to minimize the disruption of classes, and a social environment in which all students eagerly desired to participate in the study (subject trials were conducted in a mobile computer laboratory, described above), it was decided to select all students within any given class that desired to participate. No student refused. Because the computer program itself randomly assigned students to treatment conditions, it was felt that the major tests for random assignment were met by this procedure.

One hundred forty two subjects from five classes were obtained, distributed as follows:

	Class	Sequence Used	Number of Subjects
5th Grade	1	1	30
	2	3	31
	Sub-total:		61
4th Grade	1	2	31
	2	4	30
	3	5	20
	Sub-total:		81

All teachers reported wide ranges of ability within their classes. None of the students had used the computer terminals before the study although a small, unspecified number had seen the interior of the mobile van. This was the first time any of them had had an opportunity to interact with a computer.

Administration Procedures

Before students in the first two classes were asked to volunteer, they were briefly told that the purpose of the study was to see how well students responded to different kinds of incentives and different levels of problem difficulty. They were also told that they would be working arithmetic problems and that all work would be done at the computer terminals in the mobile van. Students were then asked to volunteer for the study. After the first two classes completed the prescribed procedure, it was found that only a minimum introduction was necessary. It was evident that the students were uniformly eager to participate. All introductions were given by the experimenter.

Varying numbers of students were selected from a class for each subject run, depending upon the number of available computer terminals. Because adult students were concurrently taking a college credit course at the mobile van, it was not possible to use all fifteen terminals except for two instances. The number of students "on-line" at a given time normally ranged from six to ten.

For each subject run, the experimenter took the number of students needed from their classroom and guided them to the van. Outside the van, before entering, the following instructions were given:

> When you go inside, you will be directed to an empty chair in front of a television-like screen. Sit down and put your name on the yellow card that will be in front of you. We need this because some of you may receive something with your name on it. You will then be put in touch with the computer.
>
> Please do only your own work and don't talk to your neighbor. When you are finished or if you have any questions raise your hand and we'll come to you.
>
> You will be working some arithmetic problems. Some of them will be hard, some will be easy. Do the best you can.
>
> (After the fiftieth subject the following instruction was added).
>
> Some of you will get something when you finish work and some of you won't get anything. The computer will determine what you get and will tell you before you start. So be sure to read carefully.
>
> Are there any questions before we begin?

After questions were answered, students were permitted to take their places at the terminals and their terminals were activated. The computer program directed them through the materials, as described. All questions were answered as completely as possible.

When a student wished to stop working, his terminal was deactivated and he was guided to the end of the room where he was given his reward, if any. All students were given a punched card that had the words THANK YOU punched in it. After the fiftieth subject, and when time permitted, students promised a reward were asked if they remembered what the reward was to be and why they continued to work.

Administration of the Materials at the Computer Terminal

A specific sequence of events was prescribed for each subject: 1) introduction and use of the system; 2) practice with the light pen; 3) administration of Reward Preference Inventory; 4) use of and practice with the keyboard; 5) administration of arithmetic pretest; and 6) administration of the task. A brief description of each event is presented below. Samples of the task items and a flowchart of the task programming are in Appendix F. A copy of the complete Coursewriter II program is available from the author. The following events are described in the sequence in which they were presented to the students.

Introduction and use of the system. Students were greeted with a message thanking them for their help in improving computer assisted instruction. They were then told how to progress from one display to the next.

Practice with the light pen. The use of the light pen was described. Students were first directed to look at the pen, then to

ERIC

use it in specific ways. Finally, students used it to select squares in a game of Tic-Tac-Toe with the computer.

Reward Preference Inventory. The text of the introduction and sequence of paired comparisons for the RPI is given in Appendix B. Students were given the opportunity to respond to a sample item and were given feedback verifying their choice. Twenty four pairs of questions were then presented, one pair at a time. One element of each pair was selected by means of the light pen. Students could take as long as they wished to make each choice. Once the choice was made, the next pair automatically replaced the completed one. Locations of the text and response area for each part of the pair were pseudo-randomly varied on the display screen to minimize the possibility of a response bias. No information on the results of the RPI was given the students, although the profile of each student was stored and printed later for data analysis and for the assignment of subjects to cells in the design.

Use of the keyboard. Students were instructed in the use of the numeral keys on the keyboard to prepare them to type numerical responses to arithmetic problems. A written set of instructions was used to supplement the on-line instruction to save time and to give students a permanent reference if needed.

Students were directed to find, type, and enter (submit to the computer for judging) specific numbers and the letter "n". They were then instructed in the method of correcting numbers that were incorrectly typed. Practice in the use of numbers was provided by a Tic-Tac-Toe game in which squares were identified by numbers. A square was selected by typing and entering its number.

Help was given to students needing it. It should be noted that at least half the students had difficulty with this section, but all learned the process well enough to perform adequately on the task.

Arithmetic pretest. The twenty-one item pretest was administered one item at a time via the terminal display screen and was sequenced from the easiest to the most difficult item on the basis of results from a preliminary evaluation with third, fourth, and fifth grade students. The items for the pretest were selected on the basis of an item analysis of a pool of 55 items administered to third through fifth grade students in State College, Pennsylvania. The twenty one items represented those that permitted the best discrimination when sequenced from low to high difficulty. The test was designed to be a power test such that a student would correctly answer all problems in the sequence up to his level of ability and none thereafter, and no student could solve all the problems. Problems were selected from all the basic types of arithmetic problems from addition to the subtraction of complex fractions, and at different levels of difficulty within each type, such as two, three, four, and five digit addition. The items in this pretest are reproduced in Appendix E.

The Kuder-Richardson Formula 20 reliability of the pretest, administered by the computer to 141 students, was found to be 0.762, the test mean was 5.69, the standard deviation was 3.24. The standard error of measurement was 1.58 for the 21-item test.

Sixty seconds were allowed for all single-response items; an additional twenty seconds were allowed for each additional response (such as for division with a remainder or fractions). Preliminary trial time results indicated that these times were adequate. For example, a long division problem with no remainder was allocated sixty seconds. If there was a remainder, the quotient had to be entered within sixty seconds and the student was given an additional twenty seconds to enter the remainder. A complex fraction with a whole number, a numerator, and a denominator would be allowed 100 seconds. Exceptions to this time schedule were made if the student was required to read additional directions. This occurred only for the first problem with a fractional response; five minutes were allowed to permit subjects to read how to respond to this kind of problem or to ask for help (which many of them required).

The introduction to the pretest is reprinted in Appendix E. It told the students how much time they would have for each problem and indicated that they would not be able to solve all of them. So that subjects could skip problems they could not do, an option to type an "n" was provided.

Most subjects were able to handle the mechanics of the pretest with little trouble, although most required help with the first fraction response. This latter was anticipated, but it was felt that a detailed procedure for teaching subjects to respond to fractions would be excessively time-consuming, and personal attention would insure immediate learning of the process. Some subjects had difficulty responding to the first one or two problems, but, with help, learned the process quickly.

After each problem students were told if the answer was correct or incorrect. An "n" response was followed by "O.K." The feedback was displayed for three seconds, then the next problem was automatically displayed. No summary of results was given the student when he finished, but the results were stored and tabulated for later printout and for use in the assignment of the initial difficulty level for the task. Three kinds of information were stored, the number of correct responses, the last problem correctly solved, and the number of "n" responses.

Cell assignment. Once the student completed the arithmetic pretest there was sufficient information to assign him to a cell in the experimental design; that is, he was assigned to an incentive condition and an error rate level (PCR). A flowchart detailing the decision-making process is in Appendix G. Briefly, the process was as follows. A buffer with twelve positions was established. A random number between one and twelve was generated; the resulting number was the cell and was used to set counters and switches that controlled the incentive and error rate level.

To insure an even distribution within cells, all twelve cells were filled by the pseudo-random procedure before the next set of twelve was started. The assignment to cells for each cycle of twelve was controlled by establishing a counter value equal to the largest number of subjects in a cell. As numbers were generated the new number of cases in the cell of the generated number was tested against this test value; if it was equal, the cell assignment was made. If the number of cases were greater than the test value, the generated number was rejected and another number generated and tested.

<u>The task</u>. The specific difficulty level initially assigned to a subject depended upon his performance on the arithmetic pretest and upon the percent correct range (PCR) to which he was assigned. Twenty-three levels were available, each representing a particular mix of the kind of arithmetic operation (e.g., addition, simple fraction subtraction) and difficulty within the level. These are completely described below. The assignment process was empirically designed to start the student with a kind of problem on which he would perform within the specified percentage range. The paradigms used are given in Appendix G.

The introduction to the task was presented on the display screen and consisted of the following text:

> Now I would like you to do some more arithmetic problems. They may be easy or they may be hard. You will have about a minute to do each one, so work as fast as you can.
>
> Try to answer each one. Do as many as you can. However, if you don't know how to do one, type an n like you did before.

Depending upon the assigned incentive group, one of the following messages was displayed with the above text:

(Adult approval condition)

> When you finish, I will give you a report card telling you how you did. Your grade will depend upon how well you do.

(Competition)

> When you finish, your name may be put on a chart at the entrance to show everybody how well you did. Your name will be on the chart only if you do a good job.

(Consumatory)

> If you do a good job, you will get some M & M candies. The number of M & M's you get depends upon how well you do.

(No incentive-control group)

Do as good a job as you can.

When finished reading the message, the subject pressed the space bar to start work on the task. The problems were presented in the same format used for the pretest. The same timing considerations were used, and each solution was followed by a display of the words "Right" or "Wrong" if correct or incorrect, "O.K. These are hard." if the n option was used, or, "Sorry, you took too much time." if the time limit was exceeded. Subjects paced themselves through the task by pressing the space bar each time they were ready to work the next problem. It was found that almost all subjects automatically pressed the space bar after seeing the result of their entry, thus, they worked as quickly as possible.

Twenty three combinations of problem type and difficulty (based on number of digits or complexity) were available for assignment. These combinations were ranked by difficulty on the basis of preliminary test results. The types and their difficulty rankings are given in Table 4. Sample problems for each type are in Appendix F. For the basic arithmetic operations (add, subtract, multiply, divide) the presentation format (vertical or horizontal display) was randomly selected to insure some variability within each level and to insure higher variability of performance at each level (it was found during preliminary trials that the format could make a difference in performance). This was desirable to minimize changes from level to level and to reduce the possibility that a subject would perform perfectly at one level of problem difficulty and very poorly at an adjacent level.

Table 4

Problem Types and Their Assigned Rank (Difficulty Level)

Problem Type	Rank
Addition, 2 addends, Range 1-10	1
Addition, 2 addends, Range 10-100	2
Subtraction, Range 1-200	3
Addition, 2 addends, Range 100-1000	4
Addition, 2 addends, Range 1000-10000	5
Division without remainder, Range 1-10, 1-100	6
Subtraction, Range 100-10000	7
Multiplication, Range 1-10, 1-100	8
Subtraction, Range 1000-32000	9
Simple fraction addition, like denominators, Range 1/4-7/9	10
Multiplication, Range 10-150	11
Simple fraction subtraction, like denominators, Range 1/4-8/9	12
Division with remainder, Range 2-20, 4-419	13
Multiplication, Range 50-100, 150-327	14
Division with remainder, Range 10-100, 100-10099	15
Simple fraction addition, unlike denominators, Range 1/2-13/15	16
Addition of mixed fractions, Range 1/3-9 9/10	17
Simple fraction multiplication, Range 1/2-8/9	18
Simple fraction subtraction, unlike denominators, Range 1/2-10/15	19
Multiplication of mixed fractions, Range 1/3-9 9/10	20
Simple fraction division, Range 1/2-8/9	21
Division of mixed fractions, Range 1/3-9 9/10	22
Subtraction of mixed fractions, Range 1/3-9 9/10	23

The following performance information was maintained during progress through the task:

1. Number of problems attempted
2. Number solved correctly
3. Number answered incorrectly
4. Number answered with an n
5. Number of overtimes

The rationale for permitting an "n" response was similar to that for the pretest. Students were given a time limit for their response to force them to work quickly and to prevent subjects from dawdling over their response.

The percentage of problems solved correctly was calculated after each solution. After every other problem, this percentage was compared with a maximum and minimum value of the PCR established by the random assignment. If the obtained percentage was greater than the established maximum, the difficulty level of the problem was increased by one (the subject was given more difficult problems). Easier problems were called for if the subject was performing at a level below the established minimum. Thus, although a subject started work on a given kind of problem, he might receive harder or easier problems depending upon his performance with respect to the percent correct range to which he was assigned. A typical progress of a subject who invoked this adjustment process is given in Appendix H. It is this process that insured that the task difficulty was adjusted to the individual abilities of the subjects.

To insure that students remembered the reward for which they were working and to provide them with a periodic opportunity to stop work, a message was displayed upon completion of the first nine problems and every ten problems thereafter that made reference to the reward, encouraged them to continue, but permitted them to stop if they wished. The four reward-related messages were as follows:

(Adult approval)

> If you keep working, you will get a better grade on your report card.

(Competition)

If you keep working you will be sure your name is on the board for everyone to see.

(Consumption)

You have earned three more M & M's. If you keep working, you will get more.

(No reward-control group)

Keep up the good work.

The remainder of the message instructed the subject to use the light pen to indicate that he wished to continue or stop. If he continued, an additional message told him how he could stop at any time.

Subjects were permitted to work on the task until they were ready to stop or until they had to be stopped for lunch or because school was over. Upon stopping a message was displayed thanking them for their participation. Simultaneously, data on their performance and the nature of their reward was printed on a typewriter at the front of the van (the proctor's terminal). Subjects were conducted to the front of the van where they were given their reward, if any. All subjects were given a data processing card punched with the words "THANK YOU" and were thanked for their participation. After the fiftieth subject, and when time permitted, subjects were asked if they remembered what reward they were to receive, why they worked as long as they did, why they stopped work, and how hard they thought the task was.

Data Collection

The following data were collected during the study:

a. Reward Preference Inventory:

Number of times each category was selected

ERIC

b. Arithmetic Pretest:

1. Number of problems solved correctly
2. Number of problems solved incorrectly
3. Number of n (don't know) responses
4. Number of overtimes
5. Nature of response to each item (correct, incorrect, n, overtime)

c. Task:

1. Number of problems attempted
2. Number solved correctly
3. Number solved incorrectly
4. Number answered with an n
5. Number of overtimes
6. Time spent on the task problems (measured from the time of display of the first problem of each set of ten to the entry of the solution to the last of each set)
7. Response latency for each problem (measured from the display of the character permitting an entry to the instant the response was entered to the computer).

All data collection was performed automatically by the computer and was stored in each student's performance record. In addition to the above, several other items of information were also recorded, such as the exact response and time of recording. See Chase and Bahn (1968) for a discussion of the data recording capabilities of the IBM 1500 Instructional System.

ERIC

Summary of Experimental Design

The fundamental design included an experimental group in a 3 x 4 factorial with three levels of incentive (two with expected reward, one control group with no expected reward) and four levels of PCR. The incentive levels consisted of most preferred and least preferred reward as assigned by the computer based upon each subject's RPI profile, and a control group that was administered the RPI but were not told they would receive a reward.

The number of levels of difficulty was chosen arbitrarily, but was based upon the compromise between optimizing the discriminability of the difficulty index and insuring an adequate number of subjects in each cell. An objective of 120 subjects, or ten in each cell, was established as the minimum number affording reasonable precision. The design is illustrated in Table 5. The number of subjects obtained for each cell is indicated.

Table 5

Design of the Experiment

	Mean Percent Correct Range			
Incentive	98-80	72-54	46-28	20-2
Most Preferred	n = 10	n = 13	n = 11	n = 10
Least Preferred	n = 9	n = 14	n = 8	n = 14
Control	n = 10	n = 15	n = 11	n = 10

The four levels of difficulty were defined as four ranges of performance which were maintained by the computer program during student operation. Each range centered on a target value for the percent of correct solutions. The computer continually evaluated the subject's performance and adjusted the difficulty of the problems to maintain performance within the target range selected. These ranges and their target values were:

Target Percent Correct	Range Maintained (Percent Correct)
11	2 - 20
37	28 - 46
63	54 - 72
89	80 - 98

It was decided to omit the top and bottom two percent from the range as being highly unlikely occurrences. Small unused ranges were provided between target ranges to insure some discreteness at each level. The targets and ranges were selected on the basis of the practical expectations of student performance on a task of this type.

Data Analysis

An unweighted means analysis of variance[2] was used to test the significance of the following for both dependent variables:

1. Difference between means for the four percent correct ranges
2. Differences between means for the three incentive groups.
3. The interaction between PCR and incentive.

Because the levels of both factors were determined arbitrarily, it was determined that a fixed effect AOV was the appropriate model.

---

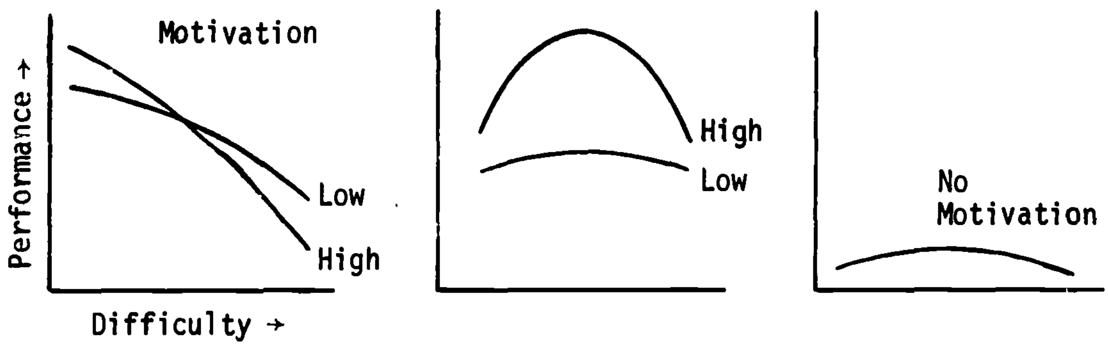
[2]The Pennsylvania State University Computer Center program ANOVUM (Analysis of Variance of Unweighted Means), prepared by Nancy Daubert, was used for these analyses.

A trend analysis was used to evaluate the linearity of the regression of problem difficulty on each dependent variable.

The implications of the results for the various theories under consideration were determined by inspection of the graphical results.

The major findings expected were the following:

1. A curvilinear relationship would be found between task difficulty and each of the dependent variables at each level of incentive. The form of the curves that might be expected from the different theories is shown in Figure 5.



a. Yerkes-Dodson and Spence-Taylor b. Atkinson c. Hunt

Fig. 5. Expected Results from the Different Theoretical Formulations.

2. The mean solution rate of the most-preferred reward group would be greater than the mean solution rate of the group receiving the least-preferred reward. Both would be greater than that for the control group.

3. The mean time spent working on problems by the most-preferred reward group would be greater than the mean time for the group receiving the least-preferred reward. Both would be longer than that for the control group.

4. An incentive by difficulty interaction would be found. It was expected that the interaction would be ordinal (no crossover) if Atkinson's Theory applied, disordinal (crossover) if the Yerkes-Dodson Law and Spence-Taylor Theory applied.

ERIC

CHAPTER IV

RESULTS

Results for Main Hypotheses

The following null hypotheses were established for each dependent variable (Task Time and adjusted Number of Attempts):

1. There is no difference between means for each percent correct range (PCR).

2. There is no difference between the means for the high incentive (HI) and low incentive (LI) groups and between the pooled incentive groups and the control (no incentive - NI) group.

3. There is no interaction between PCR and incentive.

4. The regression of each dependent variable on PCR is linear.

   a. The regression of PCR on each dependent variable at each level of incentive will be linear.

The five percent level of significance was selected for tests of significance of differences between means.

<u>Results for Hypotheses 1, 2, and 3</u>. The means, standard deviations and marginal data for each cell in the 3 x 4 factorial design are given in Tables 6 and 7. Table 6 contains Task Time data; Table 7 contains the data from the adjusted Number of Attempts. The analysis of variance summary tables used for testing the three hypotheses for each dependent variable are in Tables 8 and 9. Graphs of the data are in Figures 6 and 7. Histograms of the data, to show the distributions contributing to the marginal means, are given in

Table 6

Cell and Marginal Means and Standard Deviations for the Results for the Task Time Variable (in seconds)

Incentive Value	Difficulty (PCR) Level: Easy (98-80%)	Mod. Easy (72-54%)	Mod. Hard (46-28%)	Hard (20-2%)	Marginal Data
Most Preferred	N = 10 Mean = 3320 s.d. = 1489	N = 13 Mean = 2943 s.d. = 1762	N = 11 Mean = 2720 s.d. = 1739	N = 10 Mean = 2666 s.d. = 1886	N = 44 Mean = 2910.0 s.d. = 1685.6
Least Preferred	N = 9 Mean = 3344 s.d. = 1394	N = 14 Mean = 2092 s.d. = 1166	N = 8 Mean = 2910 s.d. = 1377	N = 14 Mean = 1837 s.d. = 1314	N = 45 Mean = 2408.6 s.d. = 1388.0
No Reward	N = 10 Mean = 3188 s.d. = 1190	N = 15 Mean = 2050 s.d. = 1751	N = 11 Mean = 2261 s.d. = 1938	N = 10 Mean = 2699 s.d. = 2360	N = 46 Mean = 2488.7 s.d. = 1839.0
Marginal Data	N = 29 Mean = 3281.8 s.d. = 1314.7	N = 42 Mean = 2340.5 s.d. = 1597.9	N = 30 Mean = 2602.1 s.d. = 1694.7	N = 34 Mean = 2334.5 s.d. = 1830.0	N = 135 Mean = 2599.3 s.d. = 1651.7

Table 7

Cell and Marginal Means and Standard Deviations
for the Results for the adjusted Number of Problems Variable

Incentive Value	Difficulty (PCR) Level Easy (98-80%)	Mod. Easy (72-54%)	Mod. Hard (46-28%)	Hard (20-2%)	Marginal Data
Most Preferred	N = 10 Mean = 158.0 s.d. = 73.0	N = 13 Mean = 79.9 s.d. = 47.2	N = 11 Mean = 58.4 s.d. = 35.6	N = 10 Mean = 52.2 s.d. = 46.1	N = 44 Mean = 86.0 s.d. = 64.5
Least Preferred	N = 9 Mean = 219.8 s.d. = 130.8	N = 14 Mean = 65.1 s.d. = 61.5	N = 8 Mean = 60.6 s.d. = 29.4	N = 14 Mean = 38.6 s.d. = 40.0	N = 45 Mean = 87.0 s.d. = 97.3
No Reward	N = 10 Mean = 177.7 s.d. = 104.0	N = 15 Mean = 53.7 s.d. = 40.7	N = 11 Mean = 52.0 s.d. = 48.3	N = 10 Mean = 38.4 s.d. = 30.3	N = 46 Mean = 76.9 s.d. = 79.6
Marginal Data	N = 29 Mean = 184.0 s.d. = 103.9	N = 42 Mean = 65.5 s.d. = 50.4	N = 30 Mean = 56.6 s.d. = 38.3	N = 34 Mean = 42.5 s.d. = 38.7	N = 135 Mean = 83.2 s.d. = 81.2

Table 8

Analysis of Variance Summary Table for Incentive and PCR, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	3853855.	2	1926928.	<1.0	0.492
2 PCR	17795732.	3	5931911.	2.199	0.092
12 Incentive x PCR	9419216.	6	1569869.	<1.0	0.744
Error	331798992.	123	2697553.		

Table 9

Analysis of Variance Summary Table for Incentive and PCR, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	5316.	2	2658.	<1.0	0.503
2 PCR	420322.	3	140107.	36.437	0.000
12 Incentive x PCR	21888.	6	3648.	<1.0	0.463
Error	472963.	123	3845.		

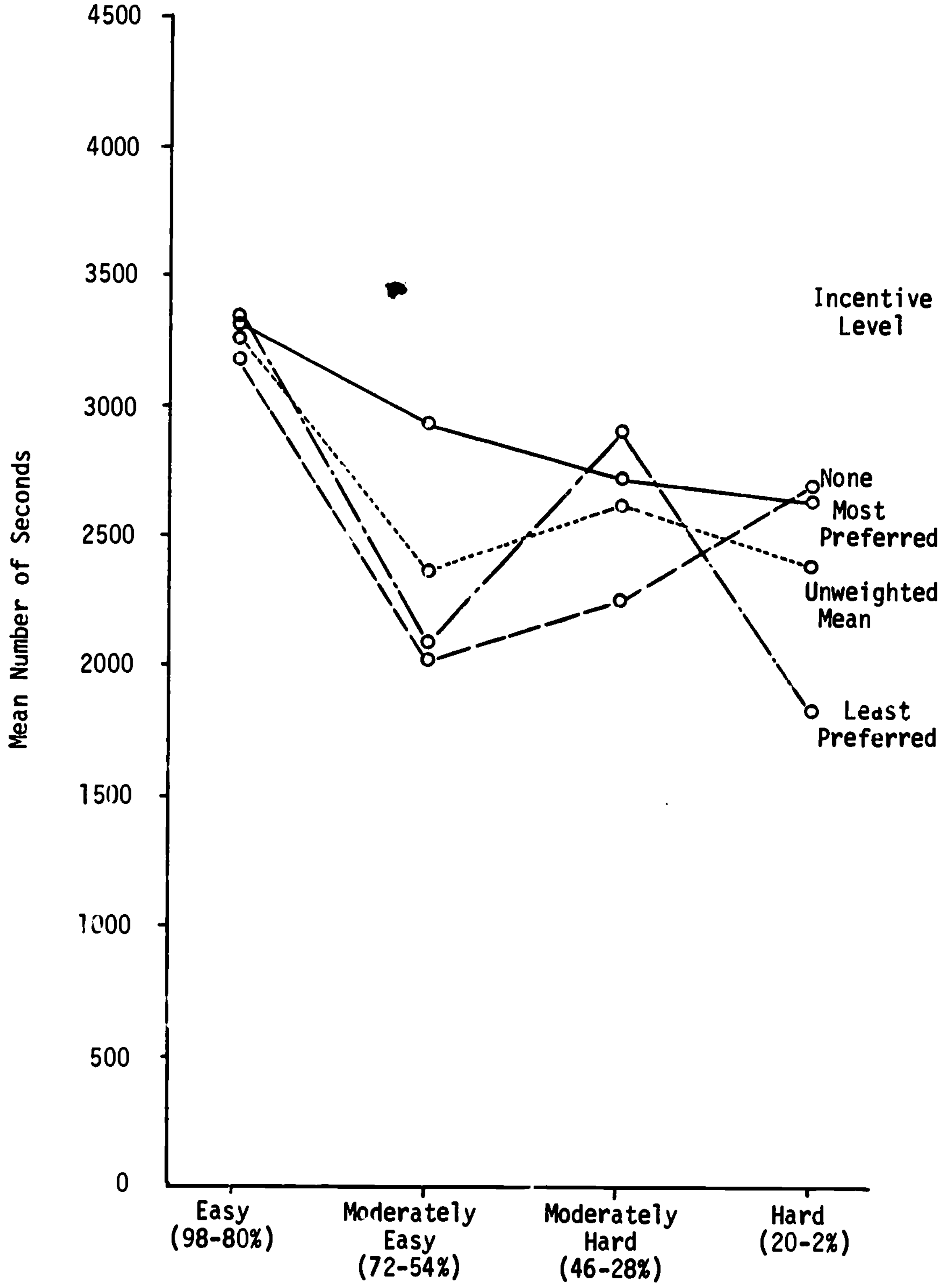


Fig. 6. Interaction between Incentive and Task Difficulty (PCR) with Task Time as the Dependent Variable.

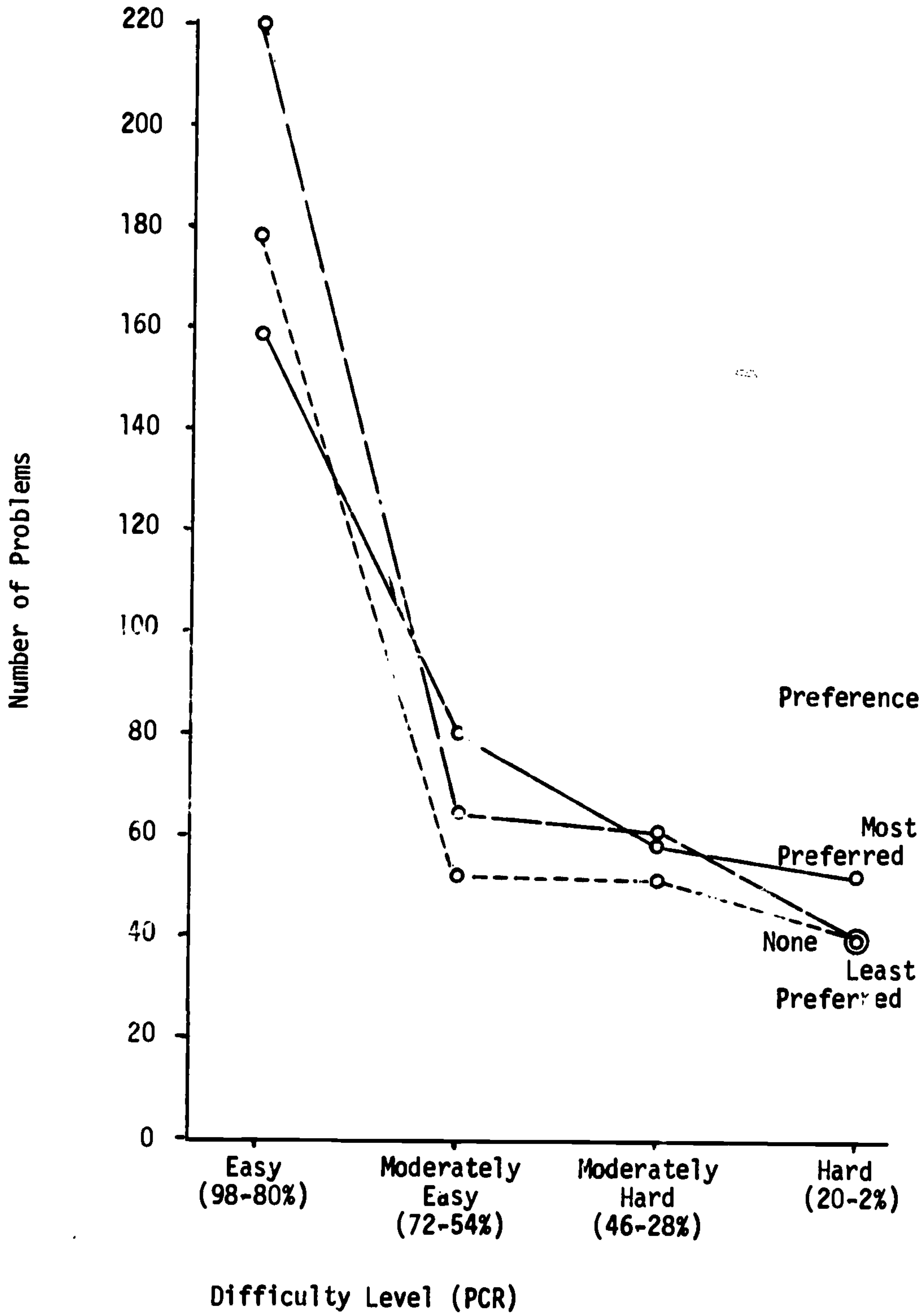


Fig. 7. Interaction between Incentive and Task Difficulty (PCR) with ANP as the Dependent Variable.

Figures 8, 9, 10 and 11. Complete data for each subject, sorted by cell, is given in Appendix I.

Homogeneity of variance was evaluated with Bartlett's test to determine if this assumption for the analysis of variance was met. The assumption of homogeneity was met for the Task Time variables but was rejected for the adjusted Number of Attempts variable (chi square = 48.6253, $p<.001$). Thus, results of analyses of the second variable must be interpreted with great caution.

Hypothesis 1. This hypothesis was concerned with differences between PCR's. For Task Time, the null hypothesis was not rejected ($F = 2.199$, $p = .092$), but it was rejected for the adjusted Number of Attempts ($F = 36.437$, $p<.001$). The reason for this difference can be seen in the tables of means (Table 6 and 7). The magnitude of the difference is much greater for the Number of Attempts than for the Time variable, although the trends are similar. Clearly, subjects worked longer on the easy (high percent correct) problems than on the other kinds, and worked many more of them. The large difference in the number worked was probably due to the very simplicity of the problems. Observation of the students and post-task interviews tended to support this view: several subjects stated that they really liked working the simple problems. Furthermore, the very ease with which they could be worked would lead to the expectation that more easy problems than hard problems could be worked in a given period of time.

The histogram data (Figures 8, 9, 10 and 11) provide further support for the observed preference for easy problems. The histograms indicate a possible linear trend, in contrast with the means; the modes for each PCR clearly decrease with an increase in difficulty. A

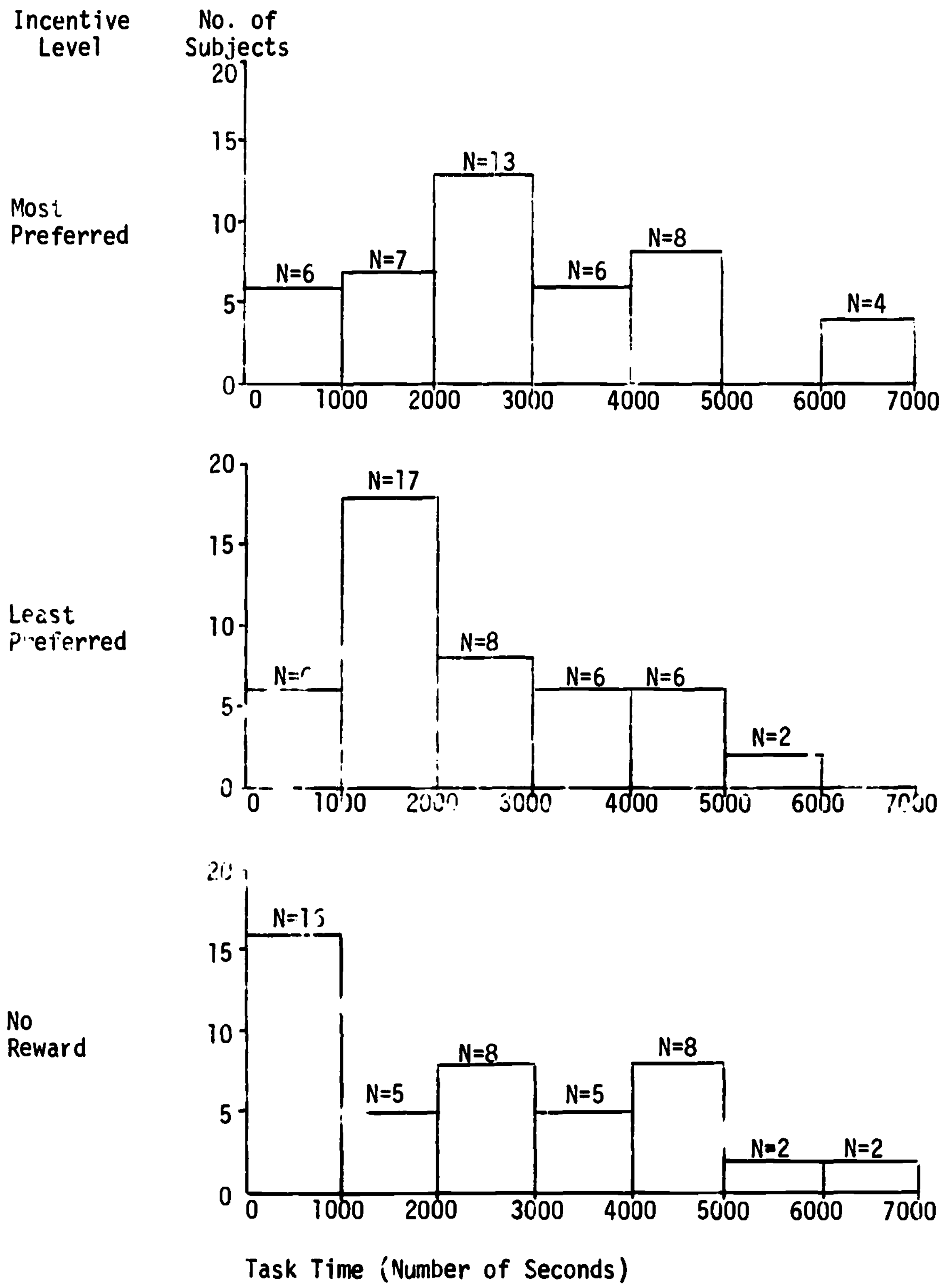


Fig. 8. Histogram Illustrating the Number of Subjects whose Task Times Fall within Specified Ranges for Each Level of Incentive.

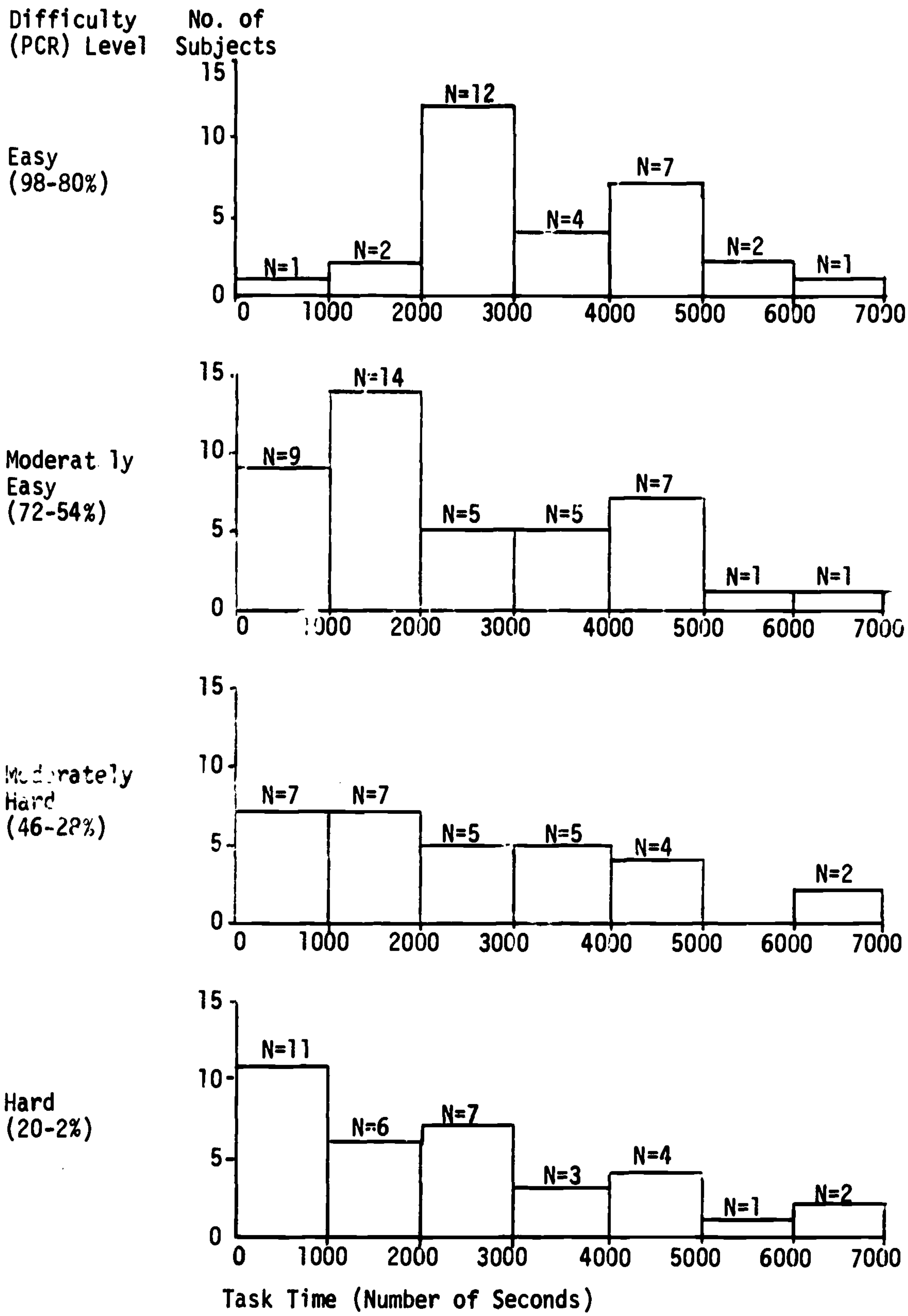


Fig. 9. Histogram of the Number of Subjects whose Task Times Fall within Specified Ranges for Each Level of Difficulty (PCR).

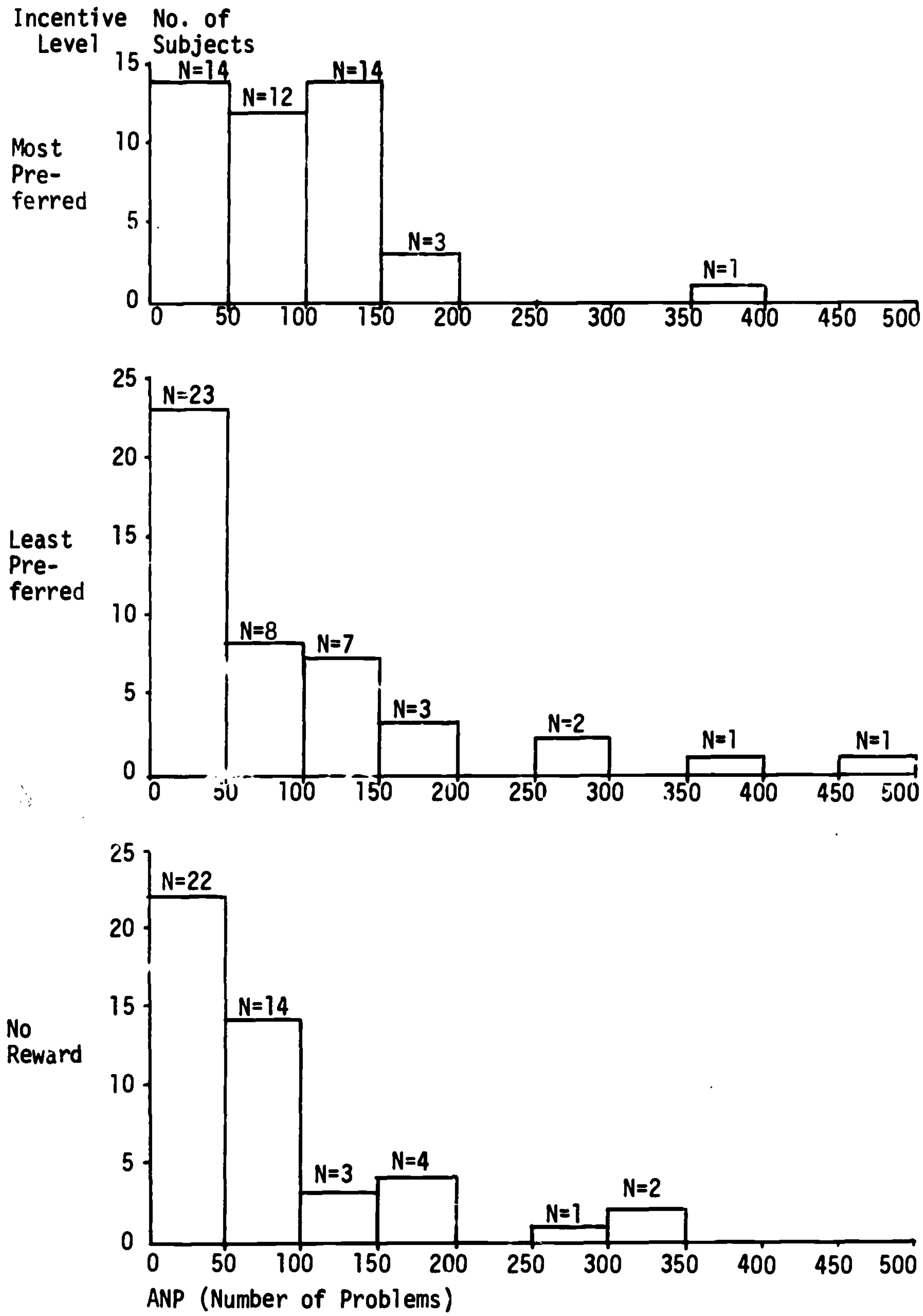


Fig. 10. Histogram of the Number of Subjects whose Adjusted Number of Problems (ANP) Fall within Specified Ranges for Each Level of Incentive.

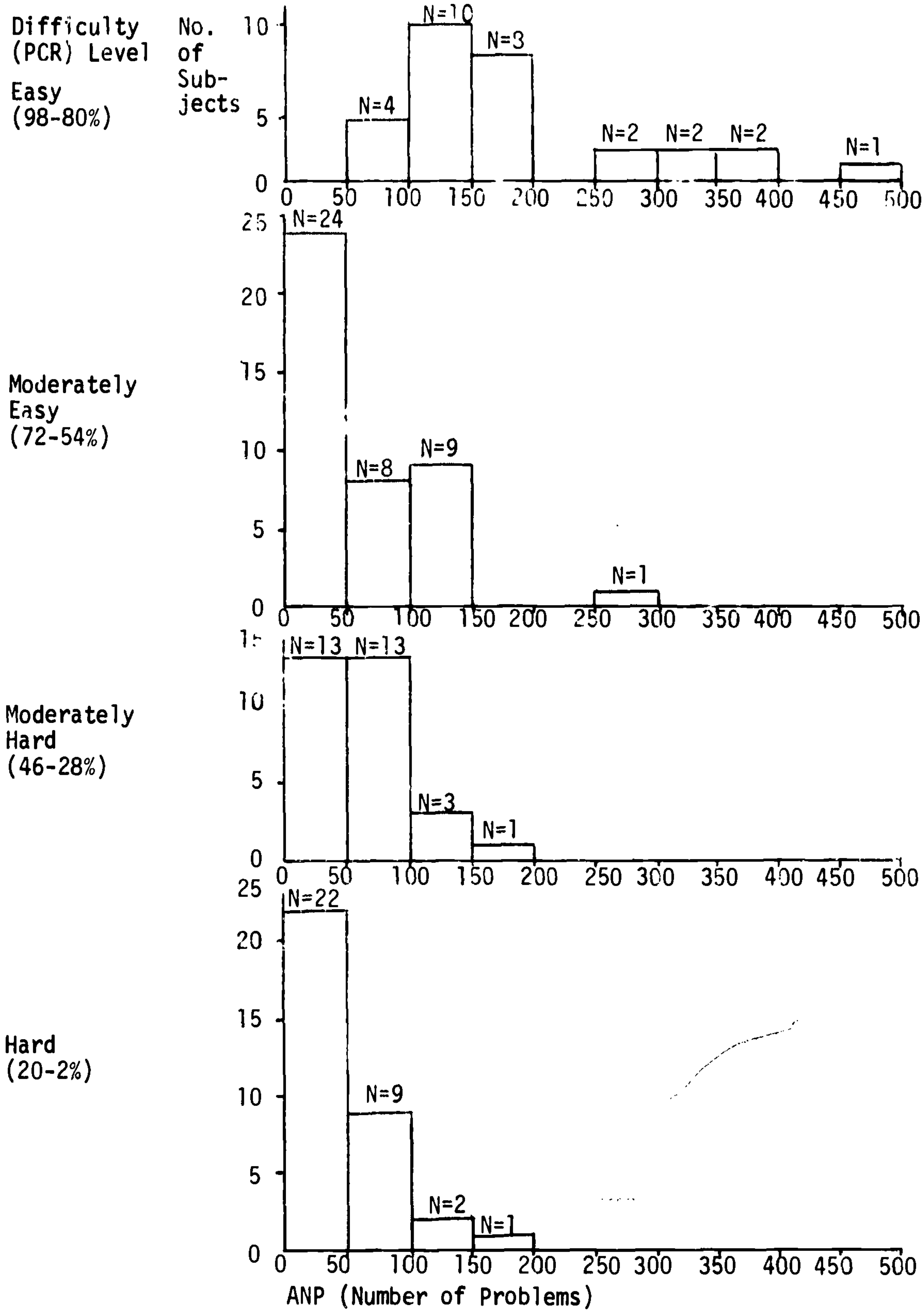


Fig. 11. Histogram of the Number of Subjects whose Adjusted Number of Problems Fall within Specified Ranges for Each Level of Difficulty (PCR).

somewhat linear relationship between difficulty and persistence would be expected, but is not reflected in the means, possibly because of the very large amount of time spent by a small number of subjects on the moderately hard and hard problems.

Hypothesis 2. Hypothesis 2 was concerned with the efficacy of the different incentive conditions. The differences for both dependent variables were clearly not significant ($p \approx .5$ for both); the null hypothesis was not rejected. Due to the lack of significance of the F tests, the pairwise comparisons were not made. The analysis indicated that the preference for incentive, or the promise of an incentive, made no difference in the performance of subjects on a persistence task in the computer terminal setting.

A review of the marginal means (Tables 6 and 7) and histograms (Figures 8, 9, 10 and 11) indicates a trend favoring the most preferred reward. The data show that on the average, subjects spent more time on the task when they expected to receive their preferred reward, and that the time spent when they expected their least preferred reward or no reward was about the same. However, subjects worked about the same number of problems whether they expected their most preferred or least preferred rewards, and those expecting a reward worked more problems on the average than those expecting no reward.

The histogram data (Figures 8, 9, 10 and 11) provide insight into the differences between the groups. The modes clearly favor the most preferred reward group. For the Task Time variable, there appears to be a linear decrease with a decrease in assigned potency of the reward, which was what was expected. The relatively high mean for the least preferred reward group appears from the histogram

(Figures 8 and 9) to be due to a small number of subjects who worked a very large number of problems, but over half the subjects worked between zero and 49 problems, indicating that the average student did not find their least preferred reward especially attractive.

Hypothesis 3. This hypothesis stated that there is no interaction between PCR and incentive. It was anticipated that an interaction would occur and its form would be similar to interactions predicted by one of the theories discussed. The interactions for both of the dependent variables failed to reach significance; the null hypothesis was retained.

A graph of the relationship between PCR and incentive for Task Time is given in Figure 6; that for the adjusted Number of Attempts is in Figure 7. To provide comparability with the curves predicted by the theories, different levels of incentive are plotted against difficulty.

The curves for the adjusted Number of Attempts appear reasonable, but there are two unexpected points on the Task Time curves: Least Preferred-Moderately Hard and No Reward-Hard. The first point (LP-MH) represents the mean of eight subjects, five with times over 3000 seconds and three with times under 2000 seconds. The second point (NR-H) represents another small cell of ten subjects, five with times below 1100 seconds and five with times above 2500 seconds (from 2600 seconds to 6700 seconds). The sharp differences between the high and low scores for each point are striking, but their validity could not be challenged.

The resulting curves bear almost no similarity to the predicted curves and, in fact, indicate almost opposite results. The Task Time curves for Most Preferred Reward and No Reward appear to be U-shaped, the opposite of the inverted U-shaped curves predicted by the theories of Atkinson and Hunt. The control, or No Reward, group was established as a specific test of the extrapolation of Hunt's theory to levels of behavior above that of the infant.

The graph of the interaction for the adjusted Number of Attempts (Figure 7) appears to resemble the predictions of the Yerkes-Dodson Law and the Spence-Taylor Theory. However, the curves for high and low motivation are the reverse of the predicted interaction: for problems of low difficulty, subjects expecting their least-preferred reward (low motivation) or no reward attempted more problems than those expecting the most preferred reward (high motivation). The most preferred reward group performed at the highest level for problems of the greatest difficulty. Thus, the results failed to produce a significant interaction and failed to support any of the theories that were expected to be applicable to the experimental conditions.

Results of Hypothesis 4. Hypothesis 4 was designed to provide further statistical evidence with respect to the theories under consideration. It was anticipated that support of Atkinson's and Hunt's theories should include regressions of PCR on the dependent variables that would have a dominant quadratic component. The results of a trend analysis for the curves in Figures 6 and 7 are in Tables 10 through 14. The computation procedures required a modification of the method suggested by Winer (1962, p. 274). The normal computational methods are designed for equal cell sizes. To adjust for the unequal cells, the product of the harmonic mean of the number of observations

Table 10

Trend Analysis for Incentive (Pooled Groups) Plotted
Against Difficulty (PCR) Level for the
Dependent Variable of Task Time.
The Graph is in Figure 6.

Component	DF	Mean Square	F	Prob.
Linear	1	9567143.1296	3.48	p>.05
Quadratic	1	3971904.5669	1.47	p>.05
Cubic	1	4715206.1741	1.75	p>.05
Error	131	2697553.		

Table 11

Trend Analysis for Most Preferred Reward Plotted
Against Difficulty (PCR) Level for the
Dependent Variable of Task Time.
The Graph is in Figure 6.

Component	DF	Mean Square	F	Prob.
Linear	1	2604019.2224	<1.0	p>.05
Quadratic	1	283771.8692	<1.0	p>.05
Cubic	1	171.2172	<1.0	p>.05
Error	40	2987047.0258		

Table 12

Trend Analysis for the Least Preferred Reward Plotted
Against Difficulty (PCR) Level for the
Dependent Variable of Task Time.
The Graph is in Figure 6.

Component	DF	Mean Square	F	Prob.
Linear	1	7232500.9091	4.30	p<.05
Quadratic	1	438189.0343	<1.0	p>.05
Cubic	1	8271753.4003	4.92	p<.05
Error	41	1681222.2036		

Table 13

Trend Analysis for the No Reward Group Plotted
Against Difficulty (PCR) Level for the
Dependent Variable of Task Time.
The Graph is in Figure 6.

Component	DF	Mean Square	F	Prob.
Linear	1	880904.2057	<1.0	p>.05
Quadratic	1	6958724.1349	2.04	p>.05
Cubic	1	703941.0700	<1.0	p>.05
Error	42	3413958.3784		

Table 14

Trend Analysis for Incentive (Pooled Groups)
Plotted Against Difficulty (PCR) Level
for the Dependent Variable of PCR.
The Graph is in Figure 7.

Component	DF	Mean Square	F	Prob.
Linear	1	310720.1102	80.81	p<.01
Quadratic	1	89921.1398	23.39	p<.01
Cubic	1	21689.5055	5.64	p<.05
Error	131	3845.		

in the columns (or cells) with the appropriate means was used rather than the column (or cell) sums for which the formulas were designed. Computational details and a rationale for the procedure are provided in Appendix J.

The resulting tables of F-ratios for the linear, quadratic, and cubic components indicate a large difference between the curves for the two variables. For the Task Time variable, no statistically significant component was found for the trend of the PCR means pooled over all levels of incentive, although the linear component approached significance and was dominant. This lack of significance may be due to the very large error mean square. Similar results for this variable were found for the Most Preferred Reward and No Reward group, although the component associated with the visually apparant shape of each curve yielded the largest F-ratio. For the Least Preferred Reward group both the linear and cubic components were statistically significant. However, the test of deviations from linearity yielded

Table 14

Trend Analysis for Incentive (Pooled Groups)
Plotted Against Difficulty (PCR) Level
for the Dependent Variable of PCR.
The Graph is in Figure 7.

Component	DF	Mean Square	F	Prob.
Linear	1	310720.1102	80.81	p<.01
Quadratic	1	89921.1398	23.39	p<.01
Cubic	1	21689.5055	5.64	p<.05
Error	131	3845.		

in the columns (or cells) with the appropriate means was used rather than the column (or cell) sums for which the formulas were designed. Computational details and a rationale for the procedure are provided in Appendix J.

The resulting tables of F-ratios for the linear, quadratic, and cubic components indicate a large difference between the curves for the two variables. For the Task Time variable, no statistically significant component was found for the trend of the PCR means pooled over all levels of incentive, although the linear component approached significance and was dominant. This lack of significance may be due to the very large error mean square. Similar results for this variable were found for the Most Preferred Reward and No Reward group, although the component associated with the visually apparant shape of each curve yielded the largest F-ratio. For the Least Preferred Reward group both the linear and cubic components were statistically significant. However, the test of deviations from linearity yielded

Table 14

Trend Analysis for Incentive (Pooled Groups)
Plotted Against Difficulty (PCR) Level
for the Dependent Variable of PCR.
The Graph is in Figure 7.

Component	DF	Mean Square	F	Prob.
Linear	1	310720.1102	80.81	p<.01
Quadratic	1	89921.1398	23.39	p<.01
Cubic	1	21689.5055	5.64	p<.05
Error	131	3845.		

in the columns (or cells) with the appropriate means was used rather than the column (or cell) sums for which the formulas were designed. Computational details and a rationale for the procedure are provided in Appendix J.

The resulting tables of F-ratios for the linear, quadratic, and cubic components indicate a large difference between the curves for the two variables. For the Task Time variable, no statistically significant component was found for the trend of the PCR means pooled over all levels of incentive, although the linear component approached significance and was dominant. This lack of significance may be due to the very large error mean square. Similar results for this variable were found for the Most Preferred Reward and No Reward group, although the component associated with the visually apparant shape of each curve yielded the largest F-ratio. For the Least Preferred Reward group both the linear and cubic components were statistically significant. However, the test of deviations from linearity yielded

a nonsignificant F-ratio ($F = 2.56$, $p > .05$). Because the form of the curve is clearly cubic, the data indicates a strong cubic trend within a primarily linear regression.

The regression of the ANP variable on PCR contains significant linear, quadratic, and cubic components. The linear component is dominant. However, the statistical significance of all three forms is most likely an artifact due to the very high value for the easy problems. Due to the similarity of the curves, the trends for each level of incentive were not calculated.

The results of the trend analyses provide further contrary evidence with respect to the Atkinson and Hunt theories. Except for the Task Time--No Reward group, the dominant trends were linear. The curve for Task Time--No Reward group had a dominant quadratic component but was the reverse of the curves predicted by the two theories. Because the No Reward condition was specifically established as a test of the extension of Hunt's Theory to higher order problem solving, the results are especially damaging.

Summary of results

a. Hypothesis 1. The hypothesis of no difference between means for each PCR was accepted for the Task Time variable but rejected for the adjusted Number of Attempts (ANP) variable.

b. Hypothesis 2. The hypothesis of no difference between incentive group means was accepted for both dependent variables. Histogram data indicate support for the predicted results.

c. Hypothesis 3. The hypothesis of no interaction between incentive and PCR was accepted for both dependent variables. Graphical support was not found for any of the theories under consideration.

d. Hypothesis 4. The hypothesis of linear regression of each dependent variable on PCR was not rejected for most levels of incentive.

CHAPTER V

DISCUSSION

Supplementary Analyses

The obtained results were clearly not congruent with the expectations in most cases. The null hypothesis could be rejected for only Hypothesis 1, and only for one dependent variable, although the other dependent variable exhibited a similar trend. It was apparent from the data that, although there existed large differences between marginal Task Time means, (at least 7 minutes, or about 12 percent between the highest mean and the next highest), the very large within cells variance resulted in a lack of statistical significance. Several *post hoc* analyses were undertaken in an attempt to determine if this variance was attributable to random variation or if its source lay elsewhere. These analyses also permitted a further evaluation of the nature of the relationships between the major variables.

The influence of sex, grade, time of day, class, and reward kind were evaluated. Due to the process of random assignment to cells, it was not possible to perform analyses of variance with more than three variables at a time without creating empty cells. Thus, it was necessary to test each of these added variables by itself against the two independent variables. In some cases, even this was not possible, forcing further division of the analyses. Due to the large number of analyses performed, the analysis of variance summary tables and graphs illustrating the data for each analysis are in Appendix K.

The sex variable. A typical assumption of most research is that there is no difference between the results for males and females. The null hypothesis of no sex difference was tested in combination with incentive and PCR using a three-way factorial analysis of variance for both dependent variables. The AOV summary tables and illustrative figures are in Appendix K, Tables 20 and 21, and Figures 15 through 18.

The null hypothesis was not rejected and the other tests yielded results identical to those previously reported. The figures indicate a systematic sex difference however: girls typically performed at a higher level than boys. This difference is dramatically indicated in Figure 15 in which the means of the girls' times were consistently eight minutes (400 seconds) or more greater than the boys' times for each level of incentive. These results are similar to those typically expected of boys and girls in a school setting, but do not agree with the typical findings of research in computer assisted instruction where sex differences are not usually found (e.g., Atkinson, 1968).

It is noteworthy that the performance by boys at the different levels of each variable was more orderly than for girls, especially at different PCR levels. The reason for this is unclear, but it might be related to the observed tendency of girls to work as a group or as buddies. This factor is discussed in greater detail below, but the abnormally high performance of the girls on moderately hard problems may be primarily attributable to three girls in the last session who seemed to be friends, each of whom worked well over an hour.

<u>The grade variable</u>. The null hypothesis of no difference in mean performance between grades was tested in a three factor analysis of variance, with incentive and PCR constituting the other two factors. The analysis of variance summary tables and illustrative figures are in Appendix K, Tables 22 and 23, and Figures 18 through 21. Bartlett's test failed to confirm homogeneity of variance for the ANP variable ($p<.0006$), thus, results for this variable should be interpreted with great caution.

The results indicated that the null hypothesis can not be rejected. However, several statistically significant interactions were obtained in addition to the established significant PCR factor for ANP. The interaction between PCR and grade was significant for both dependent variables (for Task Time, $F = 3.223$, $p \leq .025$; for adjusted Number of Problems, $F = 6.493$, $p<.001$). Plots of the means for each grade at different levels of PCR for both dependent variables yield an almost classic crossover (disordinal) interaction (Figures 18 through 21 in Appendix K). There is a clear difference in the response by the students at the two grade levels to difficulty of the content: the persistence of the fifth graders decreased in a linear fashion with an increase in difficulty on both variables. A test of this interaction of incentive with PCR for just the fifth grade yielded a highly significant PCR effect for Task Time ($F = 5.483$, $p \leq .003$), Table 24). The fourth grade students showed a marked dip in the mean amount of time spent on moderately easy problems but spent about the same mean amount of time on the other three levels. Except for the very large number of easy problems solved, a similar pattern is evident for the ANP variable.

Within the context of the study, it appears that fourth graders are less sensitive to differences in difficulty than fifth graders, which was an unexpected finding. To further evaluate this finding, an analysis of data by classes was conducted and graphed (Appendix K, Tables 24 through 27, and Figures 22 through 25). The plot of the pooled fifth grade classes provided confirmation of the linear trend. Plots of the pooled fourth grade classes also provide general confirmation of the observed trend. Thus, the results appear to be reliable for the groups studied.

The analyses of variance for the PCR and incentive variables within each grade yielded some interesting results. Although the results for the fourth grade were generally not significant for either dependent variable (excepting PCR for the ANP variable, which was expected), the results for the fifth grade produced some unique significant differences. PCR was highly significant for the Task Time variable, with a trend similar to that found for the ANP variable (as noted above). Within the ANP variable, incentive was statistically significant ($F = 3.4$, $p = .043$). A possible trend is evident in the graph (Appendix K, Figure 23), but any interpretation must be highly tentative due to the small number of subjects constituting some of the points. It seems clear that some incentive is better than no incentive, but it also appears that the attractiveness of the most preferred reward, with respect to the least preferred reward, increases with an increase in difficulty. Weak support for this observation is gained from the PCR by incentive interaction which almost reaches significance ($F = 2.162$, $p = .065$). A somewhat similar, although less clear result was found for the fourth grade class (Appendix K, Figures 24 and 25). The combination of these two led to the unexpected findings in the main analysis.

Several explanations for these findings might be posited. The different levels of maturity might have resulted in different perceptions of the task. Fifth graders may have recognized that specific PCR's existed and responded accordingly. Recall that one of the requirements for the invocation of Atkinson's Theory is the need for subjects to recognize the probabilities of success. It seems probable that the fifth grade students perceived these inherent probabilities. On the other hand, the fourth grade subjects may have failed to perceive the inherent probabilities.

The relative attractiveness of the classroom environment versus the computer terminals is an important consideration. If students found the classroom tasks aversive, the relative attractiveness of working at a computer terminal would be enhanced. If this aversion to the classroom was stronger in fourth grade students than in the fifth graders, it could account for the obtained results. Fourth graders may have preferred doing anything other than being in class, no matter how difficult the task, while fifth graders experienced a conflict of attraction from both the terminal and the classroom. In this latter case, the attractiveness of the task itself would be the factor that would determine the amount of time spent on the task. Feather (1961, 1963) considered the attraction of alternative activities to be quite important, but in this case it was impossible to measure.

The consistency of results for the limited number of classes in each grade indicates that some systematic factor is operating, but the nature of the factor may only be speculated upon. Future research in this area is clearly indicated.

The time of day variable. The time of day during which treatments are administered is frequently considered sufficiently important that an attempt is made to give all treatments during the same period. For this study, this control was not feasible. Furthermore, because of the school schedule, a longer period of time was available during the morning than during the afternoon.

The null hypothesis that was tested was: there is no difference between the mean performance of groups performing the task during the morning and during the afternoon. To evaluate the contribution of the Time of Day (TOD) factor to the error variance in the basic analysis, a three-factor analysis was performed with two dependent variables as in prior analyses. The analysis of variance summary tables and illustrative figures are in Appendix K (Tables 28 and 29, and Figures 26 through 29). Bartlett's test was significant for the ANP variable ($p<.0001$), indicating caution in the interpretation of the results for this variable.

The results for the two dependent variables are somewhat similar. The TOD variable is highly significant for Task Time results ($F = 7.564$, $p = .007$) but only approaches significance for the ANP results ($F = 2.971$, $p = .088$). The reason can be seen in the graphs (Figures 26 through 29): the Task Time curves for a.m. and p.m. are almost parallel, with more time being spent in the morning for all levels. This may be the consequence of the school schedule that provided for three hours of classes in the morning and two hours in the afternoon. The parallelness of the curves for the morning and afternoon groups, indicated statistically by low interaction F-ratios, implies that the relationship between levels of the two major dependents variables was independent of the time of day. However, the TOD difference may contribute substantially to the error variance of analyses in which it was not partialled out. Other sources of error variance are indicated by the lack of significance in the other variables.

It would be expected that the orderly difference in times would imply a similar orderly difference in the number of problems attempted. Generally this was found to be the case. As found in other analyses the PCR variable was significant, but the TOD variable failed to reach significance.

The general similarity of results for the morning and afternoon groups is significant. Students were picked in an essentially random manner for the time of day. Thus, the two times of day may be thought of as two replicates of the study. The similarity of these pseudo replicates yields some confidence in the nature of results obtained. It would have been desirable to incorporate the TOD variable into the other three-way analyses, but attempts to do this resulted in several missing cells in all cases.

Reward Preference Inventory. Central to the outcomes for Hypothesis 2 is the question of the validity of the Reward Preference Inventory. In actuality, Hypothesis 2 is a test of the validity of the RPI, but the assessment of the effectiveness of each potential incentive is important to determine the underlying reason for the results.

The mean preferences of students in the fourth and fifth grades are shown in Table 15. The ranking and magnitude of the preferences is quite similar for both grades. The order of preferences is identical to that reported by Dunn-Rankin and Shimizu (1968) and by Cartwright (1968) for several fourth through sixth grade classes. Although there were large individual deviations, the means represent the preference values of a very large number of the subjects.

The question of validity can be posed in other ways as supplementary hypotheses to hypothesis 2: 1) for each kind of incentive, there is no difference in performance whether the incentive is most or least preferred; and 2) there is no interaction between kind of incentive and preference. The results of analyses of variance (Appendix K, Tables 30 and 31) indicate that the null hypothesis for each dependent variable cannot be rejected. The graphs of the interactions (Appendix K, Figures 30 and 31) are suggestive, although one weakness with this analysis is the small number of cases represented by several points (indicated on the graph). This disparity in cell sizes is of course due to the obtained pattern of preferences. The graphs indicate that the competitive reward (name placed on a public chart) was the most effective reward, especially when selected as most preferred. The small number of subjects receiving this reward indicates that replication is required before any definite conclusions can be drawn, however. Adult approval (report card) and consumatory reward (M & M) seem to be generally equivalent in potency on the average. The

Table 15

Preferences, by Grade, for Each Kind of Incentive

Incentive	Variable	4th Grade Means (N = 81)	5th Grade Means (N = 61)	Pooled Means (N = 142)
Adult Approval (Report Card)	RPI Score	11.95	12.78	12.31
	Number MPR	21.7	24.5	22.8
	Number IPR	3.0	4.0	3.4
	Number LPR	2.3	2.0	2.2
Competitive (Name on Chart)	RPI Score	7.60	7.66	7.63
	Number MPR	1.0	2.5	1.6
	Number IPR	19.0	22.0	20.2
	Number LPR	7.0	6.0	6.6
Consumatory (M & M)	RPI Score	4.46	3.56	4.06
	Number MPR	4.3	3.5	4.0
	Number IPR	4.0	4.5	4.8
	Number LPR	17.7	22.5	19.6

Definition of variables:

RPI Score - Mean number of times selected in comparisons (Maximum = 16)

Number MPR - Number of students for which the reward was the most preferred.

Number IPR - Number of students for which the reward was the intermediate in preference.

Number LPR - Number for which the reward was the least preferred.

reversal in potency of M & M's is surprising. It would be expected that students who selected consumatory reward as their most preferred reward would work harder for them than those for which they were least preferred (the typical case). The finding that this was not the case requires explanation. Half of those receiving M & M's as their most

ERIC

preferred reward were required to solve the most difficult problems, resulting in an artifactual depression of the mean. It should be noted that within the low PCR group, those receiving M & M's as their MPR performed at a higher level than those receiving them as a LPR. Thus, the data seem to reflect an artifact of random assignment rather than a true state of affairs.

The small number of subjects receiving a report card as their LPR affords no confidence in the obtained ordering of the resulting means, although in this case the cases were evenly distributed, and included one each from the high, moderately high, and moderately low PCR. It might be expected that replication will confirm the obtained order.

The relationships of time of day and sex with the kind of reward were investigated as part of a search for confounding factors. One possibility was that there would be a markedly higher demand for M & M's before lunch than after lunch. Although the TOD variable was found to be significant with respect to Task Time ($F = 5.566$, $p = .02$), there was no mean difference in the number of problems solved to receive M & M's (see Appendix K, Tables 32 and 33, Figures 32 and 33).

For the sex by reward kind analysis, sex was significant with respect to Task Time ($F = 3.922$, $p = .05$) and almost so with respect to ANP ($F = 3.178$, $p = .077$). See Appendix K, Tables 34 and 35, and Figures 34 and 35.

On the average, girls worked much longer than boys for an A on their report card, for their name on the chart, and for no reward at all. There was no sex difference in the time spent working for M & M's. Roughly the same relationship occurred with respect to ANP.

although girls worked almost twice as many problems as boys to get their name on the chart, contributing to a significant kind of incentive effect ($F = 3.138$, $p = .028$). In contrast to the Task Time results, there was no sex difference for those receiving no reward (and variable) but girls attempted one fifth more problems to get M & M's than boys.

The use of a "paper and pencil" instrument (the Reward Preference Inventory) for the assessment of children's preferences for rewards may be questioned in that it may be a weak predictor of the amount of work a child will expend to receive a given reward. A better technique might be to give a child the opportunity to work for each reward, then, at a later date, use his performance as a basis for assigning incentives in an experimental setting. In addition to the added time involved in such a procedure, there is the added question of the stability of the effort expended for a specific reward.

Ideally, an instrument is needed that will accurately reflect a stable hierarchy of preferences for classes of rewards. The preferences should be highly related to the work expended to achieve the reward. Such an instrument should be easy to administer. The RPI has promise of being such an instrument. The prior and present uses of the RPI (Dunn-Rankin and Schimizu, 1968); Cartwright (1968); Cartwright, G. P., and Borman (personal communication); Cartwright, G. P., and Yens; and Cartwright, C. A. (personal communication) have provided some validation, although the evidence that the preference results are highly correlated with effort is not as strong as might be desired. A review of the evidence indicates that a possible weakness in the research applications of the RPI may have been the specific reward used. If all rewards had equal potency, the evidence of effectiveness might have been more clearcut.

As a result of this evidence, the RPI seemed to be a good way of obtaining the information needed for evaluating the interaction of incentive with task difficulty, especially if the attractiveness of the specific rewards offered could be maximized.

The results from this study have provided limited support for the validity of the RPI. Generally it was found that students who were promised their most preferred reward worked longer and attempted more problems than those promised their least preferred reward. This seemed to be the case for each of the rewards, although the results for M & M candies were confounded by special circumstances. The results were especially dramatic for the competitive reward but the small number of subjects indicates that replication is required before any firm conclusions can be drawn.

An informal quiz of students after completion of the task provided added information concerning the validity of the RPI. Quizzes were initiated after one quarter of the subjects had been through the task in order to gain information relating to observations by the proctors. Questioning of students was conducted only if time was available and was conducted casually. The questions were concerned with whether students remembered the reward they were to receive (if they were actually working for that reward) and why they continued work after the first nine problems (a means of determining whether the reward actually performed a motivational function).

Thirty five subjects who received a reward were quizzed. Fourfold contingency tables were constructed for each incentive by PCR cell as indicated in Table 16 below:

Table 16

Fourfold Contingency Table Indicating Frequency of MPR Subjects in Each Cell Based upon Recall of Reward and Persistence with Respect to the Mean

		Remembered correct reward?	
		Yes	No
Persistence (Mean Split)	High	8	3
	Low	2	5

The Table clearly indicates that the expected results were obtained for the sample of MPR students, but an Exact Test was not significant (based on the tables by Finney, Latscha, Bennett and Hsu, 1963). For LPR subjects, the cells were essentially balanced, a result that would be expected if the reward was not especially desirable. These results provide some further support for the validity of the RPI.

It is noteworthy that within the LPR condition those receiving M & M's almost always remembered the reward; those who did not remember almost always had the report card or name on chart as the reward. However, it may be that the informal nature of the questioning led some students to suspect that they might not get what they had been promised, thus they preferred to play "dumb." On the other hand, the periodic encouragement messages displayed on the display screen were most specific concerning M & M's in that a specific number were mentioned by each message. It may be that this message was more recallable than the rather general messages provided for the other incentives. However, the factor of detecting and remembering the reward appears to be an important one that may not have been adequately controlled in this study. Future research should more adequately insure that students know for what they are working and test for it after the students complete work.

The question of the validity of the RPI, a question critical to the results of the study, has been approached in several different ways. In each, indications of validity were found, although the evidence was not conclusive. Perhaps one reason for the lack of significant differences was the failure of students to remember the reward they were promised. Obviously, if students do not remember the reward, it will have no incentive value. Even though attempts were made to frequently remind students of the reward, it appears that the attempts were not successful in several cases. As a result, final determination of the validity of the RPI must be left to future research. However, the results of this study do indicate that the RPI has promise as a means of determining the ranking of incentives in terms of actual effort that would be expended to attain them.

<u>Additional sources of error</u>. Perhaps one of the most serious confounding factors was the tendency for students to stay and leave in small groups. Peer relationships are important at age nine and ten. Many children this age tend to do things in groups. Due to the stability of elementary classes, it would be expected that groups would form within the classroom as well as without. This problem had not been anticipated, but after observing the performance of several sets of students it appeared that small groups of students stopped

work within a short period of time. An attempt was made to counteract the group effect by seating successive students at alternate sides or alternate ends of the mobile van. However, even this did not succeed in stopping group terminations, although it may have reduced the severity of the problem.

The effects of group behavior on the results are several. Because each student was randomly assigned to a cell, each member of each group would most likely have been assigned to a different cell. Group behavior would tend to reduce differences between cell means and, in some cases, increase the within cell variance due to artifically induced extremes. Clearly, subjects who worked on a task for only as long as other members of a group would create a high error variance. These observations seem to characterize the obtained results, although the specific magnitude of the contribution of the "group effect" cannot be determined.

A second source of potential error due to the openness of the van may have been the process of awarding rewards to subjects as they finished. Because all operations had to be completed within the van, it is possible that subjects working on line saw another subject receive a reward. Or they may have known a classmate who received a specific reward in a previous session. As a result, some students may have thought they would receive a specific reward regardless of what they read on the display screen. An example of this was the response of one subject whose siblings had previously completed the task. When asked what reward she was to receive, she asked, " A THANK YOU card?" when, in reality, she was to receive a report card.

Several other sources of random error variance may have occurred due to the environment in which the students worked: 1) A small number of subjects complained of the temperature in the mobile van and may have stopped work prematurely due to this; 2) in several cases the elementary students worked concurrently with adults (including some teachers) who were taking an inservice course by computer assisted instruction. For the most part, attempts to segregate children and adults were successful (generally the students worked on one side of the van and adults on the other side). In some cases, however, it was necessary to seat a student between adults. This may have influenced the child to continue working longer than he would have without the presence of adults.

One added potential source of error may have also existed. As was determined by the RPI, adult approval is perceived as highly desirable by many students. A small number of subjects seemed to seek the attention of the project staff by asking frequent questions or looking at a staff member for approval. All questions were answered, but attempts were made to avoid showing approval or disapproval. However, the magnitude of implicit approval that subjects may have perceived is unknown, although some may have been influenced by their perceptions. Presumably these last four sources of error were randomly distributed over all conditions.

Ideally a study of this type should be conducted with subjects individually or with subjects in individual booths. Terminals were arrayed on aisles on both sides of the mobile van, so it was impossible to keep students from seeing what other students were doing. The extent of the group effect cannot be measured for the present study, but it may have been sufficient to markedly reduce the differences between conditions. Clearly, future studies should attempt to control for this variable.

Effect of imposed termination of activity

Twenty seven subjects were required to terminate their activity on the terminal prior to their voluntary completion as a result of lunch or the end of school. The subjects terminated in this way and the sum for each cell in the primary design are indicated in Appendix I (see the Force Off column). The number in each cell ranges from one to four. The marginal cell sums are not evenly distributed across conditions; for the MPR, LPR, and no reward conditions the number of imposed terminations is six, twelve, and nine respectively. This lack of balance may have biased the results in favor of the MPR condition. For the PCR variable the number of force-offs was about the same for all levels (eight, plus or minus one) except for the Moderately Hard level (46-28%) which had markedly fewer (three).

The effect of these imposed terminations was evaluated in two ways. An analysis of variance of the data with force-offs excluded yielded no difference from that with the data included. A better method was an analysis of the medians. Because the performance of all but four of the force-offs was above that of the sample mean, a medians analysis would markedly reduce the bias resulting from the imposed reduction of performance time. The cell medians, graphs, and median splits for each dependent variable are in Appendix L.

The medians data indicate that no changes in the conclusions are required. The graphs of incentive plotted against PCR are quite similar to those obtained with the means data (see Figures L1 and L2). If anything, they tend to favor the MPR condition more than did the means data, indicating that the imposed termination of subjects resulted in no significant bias in the results. The means data may have resulted in a more conservative analysis of the results with respect to the predicted outcomes.

Correlation studies

Intercorrelations were computed between several variables as a check on potential sources of systematic error and as a means of validating the selected dependent variables.

Influence of ability. The design of the study was constructed such that mathematical ability should not be a factor in the results. To check this, correlations were computed between pretest results (number correct and highest level of problem solved) and other relevant variables. Very low, nonsignificant, correlations were found between ability and the dependent variables, indicating that ability was not related to persistence in the study. A moderate but highly significant correlation ($r = .286$, $p<.101$) was found between the number of pretest problems solved correctly and a measure of the rate of problem solution (seconds per problem). This confirms the expected result that students who performed better on the pretest on the average received more difficult problems that required more time to solve.

Intercorrelations of dependent variables. The correlation between the two dependent variables, Task Time and adjusted Number of Problems, was .635, which indicates that they had about 41 percent of the variance in common. Thus, the two measures are not independent but may be measures of somewhat different persistence characteristics. Clearly, the differences in the difficulty of problems inherent in the ANP variable has no counterpart in the Task Time variable. It is believed that this correlation, although moderately high, supports the selection of two dependent variables for analysis.

The use of a computer for data collection permits a variety of dependent measures to be selected. As a check on the selected variables, intercorrelations were computed between these and other available measures; latency of response, number of responses (unadjusted), number of correct responses, and a measure of the rate of response (number of seconds to complete each problem). Table 17 shows the correlation between these variables.

The high correlations between Task Time and Latency (.950) and between ANP and the total number of responses (.931) indicate that either one of each pair could have been used. The high relationship between ANP and the number of correct answers ($r = .963$) was surprising, but is probably a reflection of the large number of correct answers generated by those with the high PCR condition. Rate seemed to be a measure of components somewhat different from the other variables. The negative correlations between Rate and measures of the number of problems solved was to be expected. Obviously, a large number of problems attempted would imply that they were solved at a rapid rate. The corresponding low correlations between Rate and time measures indicates that Rate is primarily influenced by problem difficulty. Students should be able to solve simple addition problems at a much higher rate than long division problems.

Table 17

Intercorrelations of Potential and Selected Dependent Variables

		Task Time	ANP	Latency	GNP	Correct Answers
Used	ANP	.635				
	Latency	.950	.486			
Potential	Gross No. of Responses (GNR)	.732	.931	.537		
	Correct answers	.485	.963	.336	.870	
	Rate	-.123	-.529	.054	-.549	-.539
Significance Levels, df = 133:				p≤.001		r≥.335
				p≤.01		r≥.220
				p≤.05		r≥.169

Factors in the selection of dependent variables. The moderate but significant correlation between the number of correct answers and the time variables is an added indication that students tend to work longer on easy (high PCR) problems. The low shared variance (between 10 percent and 25 percent) implies that other variables are quite important to the amount of time spent at a task. With the mass of data that can be generated by computer based experiments, the selection of appropriate dependent variables may be difficult. Many factors must be considered such as the real purpose of the study, the reliability of the information, and the actual nature of the measures.

For this study, the following observations were made with respect to potential dependent variables, although it was realized that none would be wholly satisfactory.

Task Time, although providing a good measure of persistence, would also include time spent fooling around or daydreaming (the latter was not really observed, although the former occurred in a small number of instances).

Latency would be much like Task Time, but because it would be purely a measure of the amount of time spent between the generation of the problem and the response, it might be excessively influenced by problem difficulty factors and might not accurately reflect the amount of time spent at the terminal.

Total number of problems would include the "n's" entered in an attempt to get a different kind of problem. Measures of the number of problems attempted would be unduly influenced by the PCR range (as, indeed, they were).

Adjusted number of problems (ANP - total number of problems minus the "n" responses) would be biased against subjects doing their best but simply not knowing how to do some problems and using the "n" capability. In practice, it was observed that more bias was introduced by those seeking a different kind of problem than by those using the "n" as an indication of not knowing how to do a problem.

Rate, or time per problem, would also be very sensitive to problem difficulty, but it also might tend to combine the difficulties inherent in both the time and number measures.

Based on the above findings, it is believed that the dependent measures used were the best choices for an evaluation of persistence in the context used.

Use of Relative Difficulty

One of the innovations of this study was the use of relative difficulty rather than normative difficulty as a means of assigning subjects to PCR groups. It was observed that most previous studies concerned with task difficulty either prespecified the problem content of each level or manipulated the described difficulty (the percentage the student was told he should expect to solve) but not the actual difficulty.

The use of relative difficulty, PCR, seems to have been technically successful, although the technique of maintaining subjects within a specific PCR needs improvement. Students generally seem to have been assigned to the appropriate specific difficulty level based on their pretest results. This was evaluated by measuring the spread of difficulty levels encountered by each student. (Recall that there were 23 levels of difficulty and a subject could be moved up or down the scale based upon his performance and the restrictions of the PCR to which he was assigned.) The mean spread was 5.4 levels, which meant that the average student worked problems at five adjacent levels during his time on line. Clearly, the adjustment process was active. This adjustment was necessary because students frequently could perform with high consistency at one level (e.g., they would solve correctly almost all problems at level 6 but fail most of the problems at level 7). Also, the adjustment process itself contributed to the number of levels in some cases. Once a student had worked a large

ERIC

number of problems, the adjustment reactions occurred slowly but continued until the student reached the desired success ratio.

Future applications of the PCR technique should have a better reaction to changes in performance characteristics and might include a varying mix of problems within each level. It was thought that random switching between vertical and horizontal format would produce the required variation, but it was not as successful as might have been desired. However, these are primarily refinements. The process used worked quite satisfactorily. Evidence of this is found in the low correlations between pretest results (a measure of ability) and the dependent variables (see previous section). It was the purpose of this technique to eliminate ability as a factor in the results; the data indicate that this was achieved. Thus, it appears that the procedure used in this study for assigning and maintaining ranges of difficulty is a viable technique for conducting research on task difficulty.

Summary and Implications

A summary of findings for the main hypotheses and the supplementary analyses is given in Table 18 (Chapter VII). It was hoped that the use of computer assisted instruction equipment for research would permit control over most variables that might confound the results. Although this may have been successful to some extent, the previous summary of confounding variables indicates that the effort was not entirely successful. It appears that the error variance due to these elements may have been sufficient to mask any real differences that may have existed between levels of the variables of interest. This is indicated by the fact that the marginal means and especially the histogram data reflect marked differences between groups, generally in the predicted direction, but these differences failed to reach statistical significance. However, this is one of the dangers of conducting studies in the "real world".

One of the surprising findings was the significant interaction between grade level and PCR and the fact that when the variance attributable to differences between grade was partialled out of the analysis, the incentive by PCR interaction became statistically significant. It is not known whether this grade effect was unique to the classes or school where the study was conducted or whether it might be generalizable to other schools. This can be determined only through additional research.

The significant differences between groups that worked in the morning and in the afternoon was not expected but should have been anticipated. The ranges of times and number of problems attempted were attenuated at the high end by the need to terminate several students before they themselves were ready to quit. The number of forced terminations was about the same for both periods (13 in the morning, 14 in the afternoon). How much longer the students would have worked is unknown. Both of these results occurred because of the desire to conduct the study in a somewhat naturalistic environment. However, the similar curves generated by the morning and afternoon groups instill some confidence in the results obtained.

Perhaps the greatest surprise was the lack of support for any of the theories concerned with the interaction of incentive with task difficulty. A comparison of the obtained graphs for the interaction

(Figures 6 and 7) with the predicted curves (Chapter III, page 70) illustrates the magnitude of the divergence between predicted and obtained results. It appears, in this case, that the "strong inference" technique failed to discriminate between any of the theories. The use of a dependent variable that was not directly tied to any of the theories was admittedly a risk as was the use of elementary school children in this context. Most research of this type has been done with college students. The data indicate that in this specific case the theories do not permit generalization from the specific situations to which they are tied.

The Yerkes-Dodson Law, Spence-Taylor Theory, and the predictions based on the theories of Hunt and Hebb all are tied to learning as the dependent variable. One of the basic assumptions of this study was that the same motivational elements that contribute to learning speed would also contribute to persistence and that subjects would respond in similar ways to both. Tangential supporting evidence bearing on this question was found by Atkinson and Litwin (1960). They found a very strong relationship between persistence and performance on a final examination. A second assumption made with respect to the first two approaches cited above was that incentive motivation would produce results similar to the aversive motivations called for or used to test the Yerkes-Dodson Law or Spence-Taylor Theory.

With respect to the Atkinson Theory it was assumed that high and low incentive motivation was analogous to the achievement motivation-fear of failure dichotomy, and that the actual assigned probabilities of success on a task would be perceived the same as assigned fictitious probabilities of success. The validity of these assumptions and the implications of the results for each of these approaches will be discussed separately. These discussions are based on the further assumption that the obtained results are, in general, replicable.

Yerkes-Dodson Law. Research that has supported the Law has typically used the dependent variable of the number of trials to learn an escape from a threatening environment (e.g., shock or lack of air). The data from this study indicate that no systematic relationship of the type predicted by the Yerkes-Dodson Law exists between motivation and task difficulty with respect to the dependent measures used. It was also found that for the ANP variable the strongest motivation (MPR) yielded the poorest performance on easy tasks but the best performance on the most difficult tasks which is in direct contradiction to the Law. The only similarity in results is that performance on the easy task was superior to that on other levels of difficulty. However, shock or air deprivation may permit a wider range of motivation than incentive, but these are obviously not appropriate for humans. This study supports the results of Fang (1966) and Fantino, Kasdon and Stringer (1970) in finding no confirmation for the Yerkes-Dodson Law with low stress motivation. The assumptions made concerning analogies between experimental variables for this case cannot be considered valid.

Spence-Taylor Theory. This theory is based on the relationship between anxiety and learning performance. The prediction was simple: for easy tasks subjects with high motivation would learn faster than those with low motivation; the results would be reversed for difficult tasks. Essentially, a disordinal (crossover) interaction was postulated. Only one significant incentive by PCR interaction was found (in the grade by incentive by PCR analysis for ANP); the graph of that interaction was disordinal and precisely the opposite of that predicted by the Spence-Taylor Theory. The results for the Task Time variable were less clearcut, but also offered no support to the theory. The reasons for this lack of confirmation may be several. 1) Anxiety may operate differently as a motivator than incentive; certainly it can result in more intense motivation. 2) The predictions of the theory are based on habit strenghts of competing responses during learning. This is quite different from persistence. Although generalization of the theory to non-learning situations did not seem to be ruled out due to the potential frustration component entering into the difficult tasks, the results make this assumption questionable. Its validity must be established by further research. 3) The difference in age between subjects typically used for the evaluation of the theory and those used in this study may have been a factor, although the studies of Palermo, Castaneda, and McCandless (1956) and Castaneda, Palermo, and McCandless (1956) were done with children and supported the theory.

An evaluation of the data indicated that some evidence concerning learning by subjects can be gained through an analysis of the changes in difficulty levels experienced by the students. If the initial placement of students at a specified difficulty level is in close correspondence with their actual ability level, then it might be expected that an increase in difficulty level over time would be indicative of learning or, at least, of an increase in the emergence of the correct answer as the dominant response in the response hierarchy.

Portions of the data provide support for the predictions of the Spence-Taylor theory. For easy problems there was a decrease in difficulty level for two levels of incentive (MPR and LPR), but the decrease for the low incentive condition (-1.4) was greater than for the high incentive condition (-0.2). For hard problems there was a gain in difficulty level for both groups. The gain was larger for the LPR and no incentive groups (+3.0 and +3.1 respectively) than for the MPR group (+1.8). The data are in Appendix I. These data are in accord with the theoretical expectations. However, for the easy problems the control (no reward) group posted an increase in difficulty level and the results for the intermediate ranges of difficulty were mixed. Furthermore, any interpretation of the gains or losses must assume equal intervals between the difficulty levels if comparisons are to be made, an assumption that cannot be met in this case. It appears that future research of this nature might profitably incorporate into the design some systematic evaluation of potential learning variables, but no conclusions or implications can be reasonably be drawn from the obtained data.

<u>Atkinson's Theory</u>. Because of the similarity of independent and dependent variables between those used for this theory and the present study, it was expected that if any theory received support it would be that of Atkinson. He uses probability of success as one independent variable. Feather (1961, 1963) and Maehr and Videbeck (1968) have successfully used persistence as a dependent variable. However, the motivational component of the theory, need to achieve versus need to avoid failure, was not really analogous to the traditional construct of high versus low motivation. The question was

whether high n Achievement - low n Avoid Failure was analogous to high motivation and high n Avoid Failure - low n Achievement was analogous to low motivation for the purpose of the study. The prediction of inverted U-shaped curves for both high and low motivation was based on the findings by Atkinson and Litwin (1960).

The resulting U-shaped curves for MPR and no reward (all subjects, Task Time) directly conflict with the recent findings of Maehr and Videbeck (1968), deCharms and Carpenter (1969), and the general findings by Maehr and Sjogren (1971). Although they are not in the predicted direction, they at least maintain the relationship predicted by Atkinson, convergent at the high and low difficulty ends and divergent at intermediate levels. The fourth grade Task Time data for the same levels of incentive tend to support this observation, but the lack of concurrence by the LPR group and the lack of agreement by the ANP variable cast doubt on these findings as being truly representative. However, it is also important to the theory that the students perceive task performance as an achievement situation. In spite of the instructions, some subjects may not have perceived it in this way. Thus, no clear support for Atkinson's theory can be claimed. More research is needed to clarify the obtained findings.

A critical question concerning Atkinson's theory is whether subjects perceived the intended level of probability of success assigned to the problems they worked. Due to the fluctuation of the kind of problems required to maintain their success rate within a specified PCR, it is possible that subjects failed to detect this probability. Thus, the obtained data may not have direct bearing upon the theory as defined by Atkinson.

Hunt's Theory. This study seemed to provide a good opportunity to evaluate Hunt's theory of intrinsic motivation inherent in an optimum level of task difficulty. Hebb's (1955) theory of optimum arousal seems to make similar predictions. The assumption was that the optimum level for all subjects would be somewhere between very easy and very hard. The results for the control group indicate that either this assumption is false, the use of persistence as a dependent measure is not appropriate, or the theories do not apply in this situation. For fourth grade students, optimum arousal seems to occur with the very difficult problems, but the very easy problems seem to provide the optimum arousal for fifth grade students. Thus, the optimum difficulty level may be more of an individual difference variable than had been anticipated. A possible area for future study may be the investigation of the stability of this hypothesized individual difference variable of optimum difficulty over a range of subject matters, or even within a specific subject area at different times.

CHAPTER VI

CONCLUSIONS

In spite of the failure to obtain significant differences for the main hypotheses, it is believed that this study achieved several of its objectives with respect to elementary school children. Any generalizations must be restricted to elementary age children and, due to the nature of the results, generalizations to any group other than the sample used are tenuous at best. It provided evidence, although not conclusive evidence, that the preference for a reward can make a difference in the effort expended to obtain that reward. The Reward Preference Inventory, as modified, appears to be a reasonably viable indirect means for assessing preferences for reward. Administration of the RPI, assignment of incentives, maintenance of student performance records, and feedback to the student concerning his progress toward the promised reward can be effectively accomplished by the computer in a computer assisted instruction setting. Report cards and a name on a public chart seemed to be generally effective as incentives and M & M candies were effective in some cases. The observation of Cartwright (1970) seems to still be valid: ". . . research directed toward establishing predictive validity (of the RPT) across a wide range of learning situations and for varied groups of individuals is needed (p. 157)." This study has added one more observation; more needs to be done.

Individual differences were found to be important in the assignment of percent ranges of difficulty. Based on the pooled data, it would appear that easy problems would be the level of choice in the absence of other information. Although this was clearly true for fifth grade students, it was not necessarily true in the fourth grade. Fourth graders spent an equivalent amount of time on easy, moderately hard, and hard problems. Perhaps some students liked the challenge of difficult problems. In practice, a good strategy might be to allot the first half hour of a course to a test of this preference for difficulty. The student would be able to pick ranges of difficulty at will and work in that range for as long as he desired. At the end of the allotted period the computer would evaluate the amount of time spent in each region of difficulty and select future problem difficulties based on this information and a pretest of entering ability. Such a process might optimize the interest level of the subject content for each student. It would probably be advisable to permit the student to modify his choice of difficulty level from time to time during the course of the program. Although these suggestions are primarily directed toward drill and practice applications of CAI in which immediate knowledge of results and, perhaps, corrective feedback is provided, they may also be applicable to tutorial materials presented via CAI. This study has indicated the feasibility of maintaining control over ranges of difficulty. It is felt that such procedures are practicable in most applications of CAI.

Implications for classroom teachers may also be drawn from the results of this study. It appears that procedures currently used, those of adult approval and competition (performance charts and other such devices) are effective for most students. The question of whether the obtained results are a reflection of these classroom procedures cannot be answered, but the implications are the same in any event. It cannot be denied that there appear to be some students for whom these procedures are not most effective; perhaps teachers should assess the preferences of their students and provide rewards for good performance accordingly. The superiority of the groups receiving easy problems lends support to the incremental approach to learning. Increases in task difficulty should be kept small so that the students continue to perceive the problems as relatively simple, on the average. Teachers should be aware that individual differences exist in this area also, but the effect of these differences on learning and performance are, as yet, to be specifically determined.

The results for the interaction of incentive with task difficulty are unclear. It would appear that the MPR is generally the most effective motivator, although the one significant interaction indicated that students generally worked more problems under MPR for all ranges of difficulty except the easy range for which the LPR was most effective.

The lack of clear support for any of the theories was somewhat disappointing. It appears that within the context used theories concerned with stressful motivation and learning are not transferable to positive motivation and persistence. However, this should be varified by developing a learning task which employs control over difficulty similar to those in this study. The evidence suporting the Akinson Theory is weak at best; further research focussing specifically on this theory is indicated. For example, in one approach, subjects would be specifically informed of the range of difficulty to which they are assigned. An alternative would be to specify all ranges available and let the subject choose the range he prefers at will or after every ten problems. These designs would insure conditions more like the Atkinson research than the present study provided.

This specific method was not a profitable means of evaluating Hunt's Theory. It appears that there are individual differences in the optimum level of difficulty. The design in which subjects freely select their level of preference may be a more suitable method for researching this theory.

As with much research the results of this study have posed more questions than they have answered and have pointed the way to further research. Several of these future areas of investigation have been described above. One additional major question is whether the differences between fourth and fifth grade students were real or whether they were due to some unknown external variables. One major unknown was the relative attractiveness of the classroom. In addition, several of the other uncontrolled variables, such as the "group effect" and knowledge of the correct reward, need to be controlled or manipulated. The study should be replicated with control of these variables and a means of assessing the relative attractiveness of the classroom or potential alternate activities. Because of the demonstrated effect of anxiety on performance of tasks of different difficulty (e.g., O'Neil, Spielberger and Hansen, [illegible]), some measure of anxiety should be obtained and its influence on the results of studies of this kind evaluated.

The conduct of high risk research such as described herein has the potential for leading to breakthroughs in learning, motivation, and/or behavioral theory. In most cases, the anticipated major finding does not occur, but the study leads to many other lines of profitable research. The use of a computer for research of this kind permits precise control over some areas but not over others. However, the computer can record a great deal of information that can be used for forming future hypotheses. It is hoped that the hypotheses that can be generated from this research lead to more fruitful discoveries.

CHAPTER VII

SUMMARY

The motivation of students has been a major concern of professional personnel in all fields and levels of education. However, motivation should not be considered in isolation from the task or behavior to be performed. Research has shown that there is an interaction between the difficulty of the task and the strength of motivational variables. Unfortunately, precision has been lacking in the specification of the difficulty of a task, a precision that can be gained through the application of computer control. This study investigated two aspects of motivation: the use of individual incentives to enhance learning of children and the effect of different levels of task difficulty on the effectiveness of these individualized incentives.

Three theories and one empirical law have been developed that are concerned with the motivation-by-task difficulty interaction: the theories of Spence-Taylor, Atkinson, and Hunt, and the Law of Yerkes-Dodson. The existence of these several approaches attests to the importance of the influence that task difficulty has on the effectiveness of motivational variables on learning and performance. Yerkes and Dodson (1908) were the first to describe this relationship. They found that for a more difficult task, optimum learning occurred at a lower motivational level than for an easier task. They also found a curvilinear relationship between intensity of motivation and learning. All supporting studies have used avoidance motivation.

Spence (1958) and Taylor (1956) have developed a theory based on anxiety as a drive variable. Due to the differential effect of competing response hierarchies, it was predicted that high-drive (high-anxiety) subjects would perform better than low-drive subjects on easy tasks, but the reverse would occur for difficult tasks.

Atkinson (1964) has postulated a curvilinear relationship between task difficulty (defined as probability of success) and performance, with the shape of the curve determined by a motivational factor constituted of the need to achieve and the need to avoid failure. The theory predicts that subjects with high n Achievement low n Avoid Failure will perform better if the probability of success is .5 than if it is 0.0 or 1.0. Subjects with low n Achievement - high n Avoid Failure will perform better if the probability of success is high or low that they will if it is .5.

Hunt (1961, 1965) has proposed that differences in task difficulty may create differences in intrinsic motivation. A student may feel challenged, and be more motivated, by a difficult problem than by an easy problem.

Supporting and disconfirming evidence has been found for all approaches except that of Hunt (for which almost no evidence has been found). Furthermore, these approaches led to conflicting predictions for the effect of the interacting variables upon task performance. This study was an attempt to compare all the approaches by using variables not specifically associated with any one of them. A computer was used for the control of the experimental environment to provide maximum control over the task difficulty level and data collection. It was anticipated that a study of the interaction of incentive as the motivational variable with well-controlled levels of arithmetic problem difficulty would satisfy the requirements for the comparison of the theories. It was anticipated that the results would provide information that could be used in the classroom as well as for computer assisted instruction to individualize the motivational contingencies and optimize the desire to learn or perform.

The handling of the incentive variable was a critical part of the study. Studies of the effect of incentives on performance and the efficacy of different types of incentives have yielded mixed results. However, the Reward Preference Inventory (RPI) of Dunn-Rankin and Shimizu (1968) has indicated promise of being a valid instrument for determining the preference ranking of classes of rewards by children. A modified version of the RPI was selected as the means for assessing individual preferences for rewards that would serve as incentives. It was assumed that a child would perform better if upon completion of the task, he expected that he would receive a reward that was highly desirable to him. The modified RPI afforded discrimination between three classes of rewards: adult approval, competition, and consumable.

Several outcomes were specified for the research: 1) it would provide evidence that would permit the comparision of the theories described above with respect to persistence in problem solving; 2) information on the effectiveness of measuring the preference for incentive through the medium of computer assisted instruction would be provided; 3) the effectiveness of individualizing rewards for elementary school children would be evaluated. It was also anticipated that the results might provide a basis for developing individualized incentive contingencies for students learning via computer assisted instruction and would provide information concerning the specification of optimum individualized levels of difficulty for drill and practice tasks.

The study was designed as a persistence study due to the sensitivity of persistence measures to subtle differences in motivation. Two measures of persistence were selected as dependent variables: Task Time (the actual duration of time spent working on the task itself) and adjusted Number of Problems attempted (ANP, a measure of the actual effort expended in terms of the number of problems a student tried to solve). The independent variables were Incentive and Task Difficulty. The basic experimental design included a three by four factorial with three levels of incentive (the promise of the most preferred or least preferred reward, based on the RPI results, and a control group that was promised no reward), and four levels of difficulty (four ranges of percent correct responses [PCR], specifically, 98-80%, 72-54%, 46-28%, and 20-2% correct).

The study was conducted using an IBM 1500 Instructional System located in a mobile van. One hundred-forty-two fourth and fifth grade students worked at terminals that contained a keyboard and a cathode ray tube (CRT) display screen. The students first learned to use the terminal, then they completed the Reward Preference Inventory and an arithmetic pretest. Each student was randomly assigned by the computer to one of the twelve cells of the experiment. The nature of the reward was determined by the cell assignment and the RPI results and was displayed to the students prior to work on the task. The three specific rewards used were a report card with an A grade (adult approval), name placed on a public list of students who did outstanding work (competition), and M & M candies (consumable).

The arithmetic persistence task itself consisted of randomly generated arithmetic problems. Twenty-three different kinds of problem could be generated by the computer with each kind representing a different level of difficulty. The initial level of difficulty was determined by an analysis of the results of the arithmetic pretest and the assigned cell. After the initial level, the subject's performance was continuously evaluated and compared with the prescribed limits for the cell. Deviations from the prescribed limits produced adjustments in the difficulty level so that problem difficulty was maintained relative to the individual, an absolute difficulty, rather than relative to the group which is a normative difficulty. The student was permitted to work problems until he was ready to stop (in a few cases, students had to be stopped for lunch or because school had ended). A message describing the promised reward (or a message of encouragement if none was promised) and an option to continue work or stop was displayed after every ten problems. When a student stopped he was given the promised reward, asked about his attitudes toward the process, thanked for his participation, and returned to class.

All experimental conditions were controlled by the computer and all data were recorded by the computer except for the attitude data.

Four major hypotheses and several supplementary hypotheses were specified for the study. These are briefly stated in Table 18 with a summary of the results for each hypothesis. Unweighted means analyses of variance were used to test most hypotheses. The number of subjects used for the final analyses was 135.

Table 18

Summary of Hypotheses and Results

Hypothesis	Results
Main Hypotheses:	
Hypothesis 1. There is no difference between means for each range of percent correct (PCR); e.g., between ranges of difficulty.	For Task Time, no differences were found. For ANP there was a significant difference between the Percent Correct Ranges($p<.001$).
Hypothesis 2. There is no difference between the high and low incentive groups and between the pooled incentive groups and the no incentive (control) groups.	No significant difference was found for either dependent variable, but histogram data indicate some support for the predicted differences (most preferred>least preferred>no incentive).
Hypothesis 3. There is no interaction between PCR and incentive.	No interaction was found for either dependent variable. Graphical analysis yielded no support for any of the theories under consideration.
Hypothesis 4. The regressions of the pooled and individual levels of incentive on PCR will be linear.	Linear regressions were dominant in all cases.
Supplementary Hypotheses:	
Hypothesis A. There is no difference due to sex.	No difference was found for either dependent variable, but girls' performance was typically better than that of boys.

Table 18 (Continued)

Hypothesis	Results
Supplementary Hypotheses (Continued):	
Hypothesis B. There is no difference due to grade.	No difference was found for either dependent variable, but several interesting effects were found: 1. A disordinal interaction between PCR and grade was significant for both dependent variables ($p<.025$). 2. For the adjusted Number of Problems variable significant interactions were found between Incentive and PCR ($p = .03$) and between Incentive and Grade ($p<.001$), and significance was approached by Incentive ($p = .054$) and the Incentive by Grade interaction ($p = .058$).
Hypothesis C. There is no difference due to time of day.	For Task Time, the difference between morning and afternoon was significant ($p<.01$); students worked longer in the morning than in the afternoon. For ANP, the difference was not significant. For both dependent variables, curves of the independent variables plotted for both times of day were essentially parallel.
Hypothesis D. There is no interaction between kind of incentive and preference for the incentive.	No interaction or differences in simple effects were found for either dependent variable.

Table 18 (Continued)

Hypothesis	Results
Supplementary Hypotheses (Continued):	
Hypothesis E. There is no interaction between kind of incentive and the time of day.	No differences were found.
Hypothesis F. There is no difference between kind of incentive and sex.	For ANP, a significant difference was found between Kinds of Incentive (p = .028). The competitive reward yielded the most problems attempted for both sexes.

The results generally were not congruent with the expectations. However, the marginal means and histograms were in the expected direction for the Incentive variable. It was felt that the lack of statistical significance was due to the very large error term. Supplementary analyses tended to support this observation in that time of day, class, and, to some extent, sex may have confounded the main analyses. Thus, there is some indication that the promise of a reward will cause students to work longer on a task in a computer assisted instruction setting than no promise of a reward, and the promise of a most preferred reward will be more potent than the promise of a reward of lesser preference.

The major surprise was the lack of support for any of the theories of interaction between motivation and task difficulty. No interaction was found and the curvilinearity predicted by the Atkinson, Yerkes-Dodson, and Hunt approaches was absent. That curvilinearity which was apparent in graphs was the reverse of that predicted. A reversal (with the ANP variable) was also found with respect to the Spence-Taylor Theory: for easy problems, the highest motivation yielded the poorest performance, but this relationship was reversed for the hard problems. This divergence from the anticipated results may have been due to the lack of generalization from the specific variables (independent or dependent) assumed by theoretical formulations, to the differences in the experimental design, the differences in the medium of the studies, or the differences in the subject populations. Future research might profitably be oriented toward employing the incentive-task difficulty paradigm with modifications to make it more like that required by each individual theory.

One additional surprising finding that emerged from the study was the significant interaction between grade and PCR. There was a clear difference in the response to differences in difficulty by the two grade levels. For the sample used, this seemed to be a reliable occurance and certainly forms the basis for future research. The differences may have been due to the relative attractiveness of the computer terminals and classroom, the perception of difficulty of the task itself or other unsuspected factors.

The Reward Preference Inventory performed its task of discriminating preferences for rewards reasonably well. This study provided additional evidence supporting its validity as a research instrument. The modifications and the specific incentives used seem to have worked reasonably well. For the report card and the name on the chart subjects performed better when these were the most preferred reward than when they were preferred least. There was a reversal for the M & M candies, but this may have been an artifact of the small number receiving candies as their most preferred reward.

The computer assisted instruction technique seems to be an excellent method for conducting research of this kind. The desired control over the major variables was obtained. The random generation process for producing problems of different kinds worked well and may be a good method for generating problems for classroom practice. Further research might be directed toward giving students control of the difficulty level so that individual differences between students could be more fully investigated. Several other areas for research have also emerged from the study. The manipulation of incentives and task difficulty seems to be a worthwhile topic for research in anticipation of the future increase in the automation of learning.

REFERENCES

Aiken, L. R., Jr. Attitudes toward mathematics. Review of Educational Research, 1970, 40, 551-596.

Atkinson, J. W. An introduction to motivation. Princeton: Van Nostrand, 1964.

Atkinson, J. W., & Litwin, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. Journal of Abnormal and Social Psychology, 1960, 60 52-63.

Atkinson, R. C. Computerized instruction and the learning process. American Psychologist, 1968, 23, 225-239.

Birney, R. C., Burdick, H., & Teevan, R. C. Fear of failure. New York: Van Nostrand-Rinehold, 1969.

Blake, K. A. Studies of the effects of systematic variations of certain conditions related to learning, III. Task conditions. The University of Georgia, Athens, Georgia, Cooperative Research Project No. 1356, 1964.

Breiling, J. P., Shipman, H., Milligan, J., & Pepin, L. Clugies, snirkles, and models, three systems of token reinforcement in the grade school classroom. Paper presented at American Educational Research Association meeting, Minneapolis, March, 1970.

Broadhurst, P. L. The interaction of task difficulty and motivation: The Yerkes-Dodson law revived. Acta Psychologica, 1959, 15, 321-338.

Brown, J. S. The motivation of behavior. New York: McGraw-Hill, 1961.

Cartwright, C. A. B. The efficacy of preferential incentives with elementary school children. Unpublished doctoral dissertation, Pittsburgh, Pa., University of Pittsburgh, 1968.

Cartwright, C. A. Efficacy of preferential incentives with elementary school children. Journal of Educational Psychology, 1970, 61 152-158.

Cartwright, C. A., & Cartwright, G. P. Determining the motivational systems of individual children. Teaching Exceptional Children, Spring, 1970, 143-149.

Castaneda, A., Palermo, D. S., & McCandless, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. Child Development, 1956, 27, 327-332.

Chase, F. N. & Bahn, T. A. Analysis of student performance records. In H. E. Mitzel, Experimentation with computer-assisted instruction in technical education. Semi-Annual Progress Report. University Park, Pa.: Computer Assisted Instruction Laboratory, The Pennsylvania State University, Report No. R-18, 1968.

Child, I. L. Personality. Annual Review of Psychology, 1954, 5, 149-170.

Clifford, M. M. Motivational effects of competition and goal setting in reward and non-reward conditions. The Journal of Experimental Education, 1971, 39, (#3), 11-16.

Cole, L. W. The relation of strength of stimulation to rate of learning in the chick. Journal of Animal Behavior, 1911, 1, 111-124.

Confer, R. Effectiveness of CAI for general mathematics repeaters: Some concerns. Paper presented at the meeting of the Association for the Development of Instructional Systems, Columbus, Ohio, August 18-20, 1971.

de Bonis, M. Niveau de motivation discriminabilite des responses et performance. Annee Psychologique, 1967, 67, 109-118.

de Charms, R. & Carpenter, V. Measuring motivation in culturally disadvantaged school children. In H. J. Klausmeier and G. T. O'Hearn (Eds.), Research and development toward the improvement of education. Madison, Wisconsin: Dembar Educational Services, 1968.

Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation in the kitten. Journal of Animal Behavior, 1915, 5, 330-336.

Dunn-Rankin, P. & Shimizu, M. Reward preference patterns in elementary school children. Paper presented at the Annual Meeting of the Hawaii Psychological Association, Honolulu, Hawaii, May 3, 1968.

Dunn-Rankin, P. & Shimizu, M. Reward preference patterns in elementary school children. International Journal of Educational Sciences, 1969, 3, 53-62.

Entin, E. E. The relationship of the theory of achievement motivation and performance on a simple and complex task. Dissertation Abstracts, 1968, 29 (3-B), 1160-1161.

Eysenck, H. J. Experiments in motivation. New York: Macmillan Co., 1964.

Eysenck, H. J. & Gillan, P. W. Speed and accuracy in mirror drawing as a function of drive. In H. J. Eysenck, Experiments in motivation. New York: Macmillan Co., 1964.

Eysenck, H. J. & Warwick, K. M. The effects of drive level on a multiple-choice reaction task. In H. J. Eysenck, Experiments in motivation. New York: Macmillan Co., 1964.

Fang, M.C.S. Effect of incentive and complexity on performance of students from two social class backgrounds on a concept identification task. Research and Development Center for Learning and Re-Education, The University of Wisconsin, Madison, Wisconsin, 1966.

Fantino, E., Kasdon, D., & Stringer, N. The Yerkes-Dodson law and alimentary motivation. Canadian Journal of Psychology, 1970, 24, 77-84.

Feather, N. T. The relationship of persistence at a task to expectation of success and achievement-related motive. Journal of Abnormal and Social Psychology, 1961, 63, 552-561.

Feather, N. T. Persistence at a difficult task of intermediate difficulty. Journal of Abnormal and Social Psychology, 1963, 66, 604-609.

Feldman, M. P. Motivation and task performance; a review of the literature. In H. J. Eysenck, Experiments in motivation. New York: Macmillan Co., 1964a.

Feldman, M. P. Response reversal performance as a function of drive level. In H. J. Eysenck, Experiments in motivation. New York: Macmillan Co., 1964b.

Finney, D. J., Latcha, R., Bennett, B. M., & Hsu, P. Tables for testing significance in a 2 x 2 contingency table. Cambridge, England: Cambridge University Press, 1963.

Goodyear, F. H. An experimental study of the motivational effect of punishment and reward anticipation on the listening comprehension of college students. Dissertation Abstracts International, 1970, 30 (12-A), 5551.

Goyen, J. D. & Lyle, J. G. Effect of incentives and age on the visual recognition of retarded readers. Journal of Experimental Child Psychology, 1971a, 11, 266-273.

Goyen, J. D. & Lyle, J. G. Effect of incentives upon retarded and normal readers on a visual-associate learning task. Journal of Experimental Child Psychology, 1971b, 11, 274-280.

Haaf, R. A., Feldstein, J. H., and Witryol, S. L. A developmental study of children's incentive-object preferences. Developmental Psychology, 1970, 3 (2), 275.

Haaf, R. A. Cue and incentive motivational properties of reinforcers in children's discrimination learning. Journal of Experimental Child Psychology, 1971, 11, 1-10.

Hall, B. W. Anxiety, stress, task difficulty and achievement via programed instruction. Paper presented at the American Educational Research Association meeting, Minneapolis, March, 1970.

Hammes, J. A. Visual discrimination learning as a function of shock-fear and task difficulty. Journal of Comparative and Physiological Psychology, 1956, 49, 481-484.

Hebb, D. O. Drives and the C. N. S. Psychological Review, 1955, 62, 243-254.

Hilgard, E. R. & Bower, G. H. Theories of learning, third edition. New York: Appleton-Century-Crofts, 1966.

Hull, C. L. A behavior system: An introduction to behavior theory concerning the individual organism. New Haven: Yale University Press, 1952.

Hunt, J. McV. Intelligence and experience. New York: Ronald Press Co., 1961.

Hunt, J. McV. Intrinsic motivation and its role in psychological development. In D. Levine, Nebraska Symposium on Motivation. University of Nebraska Press, 1965.

Hunt, J. McV. Using intrinsic motivation to teach young children. Educational Technology, 1971, 11 (2), 78-80.

Karabenick, S. A. & Youssef, Z. I. Performance as a function of achievement motive level and perceived difficulty. Journal of Personality and Social Psychology, 1968, 10, 414-419.

Kight, H. R. & Sassenrath, J. M. Relation of achievement motivation and test anxiety to performance in programmed instruction. Journal of Educational Psychology, 1966, 57, 14-17.

Kimble, G. A. Hilgard and Marquis' conditioning and learning. New York: Appleton-Century-Crofts, Inc., 1961.

Lackner, L. S. Reinforcer effectiveness and social class. Paper presented at the American Educational Research Association meeting, Minneapolis, March, 1970.

Lawrence, D. H. Learning. In P. R. Farnsworth & Q. McNemar (Eds.). Annual Review of Psychology, 9. Palo Alto: Annual Reviews, Inc., 1958.

Li, J. C. R. Statistical inference II. The multiple regression and its ramifications. Ann Arbor, Michigan: Edwards Bros., Inc., 1964.

Maehr, M. L. & Sjogren, D. D. Atkinson's theory of achievement motivation: First step toward a theory of academic motivation? Review of Educational Research, 1971, 41, 143-161.

Maehr, M. L. & Videbeck, R. Predisposition to risk and persistence under varying reinforcement-success schedules. Journal of Personality and Social Psychology, 1968, 9, 96-100.

Mitzel, H. E., Hall, K. A., Suydam, M. N., Jansson, L. C., & Igo, R. V. A Commonwealth consortium to develop, implement and evaluate a pilot program of computer-assisted instruction for urban high schools. University Park, Pa.: The Pennsylvania State University, Computer Assisted Instruction Laboratory, Final Report No. R-47, July, 1971.

Noffsinger, T. The effects of reward and level of aspiration on an educational task with children identified by a problem behavior checklist. Mimeo, undated.

O'Leary, K. D. & Drabman, R. Token reinforcement programs in the classroom: A review. Psychological Bulletin, 1971, 75, 379-398.

O'Neil, H. F., Spielberger, C. D., & Hansen, D. N. Effects of state anxiety and task difficulty on computer assisted learning. Journal of Educational Psychology, 1969, 60, 343-350.

Palermo, D. S., Castaneda, A., & McCandless, B. R. The relationship of anxiety in children to performance in a complex learning task. Child Development, 1956, 27, 333-337.

Platt, J. R. Strong inference. Science, 1964, 146, 347-353.

Shrable, K., & Sassenrath, J. M. Effects of achievement motivation and test anxiety on performance in programmed instruction. American Educational Research Journal, 1970, 7, 209-220.

Skinner, B. F. Why teachers fail. Saturday Review, October 16, 1965, 80 ff.

Spence, K. W. Behavior theory and conditioning. New Haven, Conn.: Yale University Press, 1956.

Spence, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. American Psychologist, 1958, 13, 131-141.

Spielberger, C. D. The effects of anxiety on computer assisted learning. Tallahassee, Fla.: Florida State University, Computer-Assisted Instruction Center, Technical Report No. 7, June 15, 1970.

Standish, R. R., & Champion, R. A. Task difficulty and drive in verbal learning. Journal of Experimental Psychology, 1960, 59, 361-365.

Stennett, R. G. The relationship of performance level to the level of arousal. Journal of Experimental Psychology, 1957, 54, 54-61.

Suppes, P., Jerman, M., & Brian, D. Computer-assisted instruction: Stanford's 1965-66 arithmetic program. New York: Academic Press, 1968.

Taylor, J. A. Drive theory and manifest anxiety. Psychological Bulletin, 1956, 53, 303-320.

Travers, R. M. W. Reinforcement, a review of selected research. Final report. September, 1963, USHEW-OE/University of Utah, Bureau of Educational Research.

Ulrich, R. E., Louisell, S. E., & Wolfe, M. The learning village: A behavioral approach to early education. Educational Technology, 1971, 11 (2), 32-45.

Wickens, D. D. The effect of competition on performance of tasks of differing degrees of difficulty. Psychological Bulletin, 1942, 39, 595.

Willett, R. A. Experimentally induced drive and performance on a five choice serial reaction task. In H. J. Eysenck, Experiments in Motivation. New York: Macmillan Company, 1964a.

Willett, R. A. Situation-induced drive and paired-associate learning. In H. J. Eysenck, Experiments in Motivation. New York: Macmillan Company, 1964b.

Wine, J. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76 (2), 92-104.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wrobel, P., & Resnick, L. B. An investigation of the effects of the application and removal of a token economy on the working behavior of sixteen headstart children. Paper presented at the American Educational Research Association meetings, Minneapolis, March, 1970.

Yerkes, R. M., & Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurological Psychology, 1908, 18, 459-482.

APPENDIX A

Definition of Terms and Abbreviations

ERIC
Full Text Provided by ERIC

DEFINITION OF TERMS AND ABBREVIATIONS

It is recognized that different theorists may define common terms somewhat differently. For example, "motivation" and "drive" are frequently used interchangeable (e.g., Hull, Atkinson and Eysenck) as constructs that refer to a nonspecific energizing function. Furthermore, motivation has been defined quite differently by different theorists (see Atkinson, 1964, and Hilgard and Bower, 1966). For the purpose of this paper, a variable may be thought of as motivational if: 1) it tends to facilitate of energize responses; 2) its termination or reduction following a new response leads to the learning of that response; 3) sudden increases in the strength of the variable lead to the abandonment of several responses in the response hierarchy; and 4) its effects on behavior cannot be attributed to other processes such as learning, sensation, innate capacities, or sets (after Brown, 1961). This study is primarily concerned with the first of these aspects of motivation.

"Reinforcement" has also been defined in a variety of ways. A good empirical definition is provided by Kimble (1961): "Reinforcement refers to the occurrence of one of a certain class of events called reinforcers, in the proper relation to the to-be-learned response; the proper relation being that which tends to increase the probability that the response so reinforced will occur with representation of the situation. The failure to reinforce a response decreases the probability of occurrence of the response," (p. 5).

A "reward" is a pleasant stimulus that is a positive reinforcer. Cartwright (1970) defines a reward as a stimulus, given to a subject after the occurrence of a particular response in a given situation, which tends to increase the frequency of the response. "Incentive" is defined as referring to the expectation that a particular reward will be obtained after behaving in a certain way. However, a reward may also be given after a sequence of particular responses.

GLOSSERY OF ABBREVIATIONS

ANP - Adjusted Number of Problems (Dependent Variable)
CAI - Computer Assisted Instruction
GNR - Gross Number of Responses
H - Hard Problem range (20-2% correct)
HI - High Incentive
HiD - High Drive condition
Inc.- Incentive (Independent Variable)
IPR - Intermediate Preference for a Reward
LI - Low Incentive
LoD - Low Drive condition
LP - Least Preferred
LPR - Least Preferred Reward
MAS - Manifest Anxiety Scale
MH - Moderately Hard range (46-28% correct)
MPR - Most Preferred Reward
NI - No Incentive condition
NR - No Reward
PCR - Percent Correct Range (Independent Variable)
RPI - Reward Preference Inventory
TAQ - Test Anxiety Questionnaire
TAS - Test Anxiety Scale
TAT - Thematic Apperception Test
TOD - Time of Day

APPENDIX B

Reward Preference Inventory

REWARD PREFERENCE INVENTORY

Pilot Studies and Modifications

In order to evaluate the usefulness of the RPI in a CAI environment, a series of pilot studies has been conducted under the direction of G. P. Cartwright that replicates in a CAI setting the study by C. Cartwright (1970). The independent variable was the promise of the most preferred or least preferred reward; the dependent variables were time spent on a repetitive task of finding 5's and the number of 5's found.

The first pilot study, with about thirty children aged eight to ten, led to the following conclusions concerning the use of the RPI in a CAI environment: 1) it is difficult to operationally provide effective peer approval or independence rewards; 2) children tend to forget the reward they were promised; and 3) the operation of a CAI terminal is, in itself, a strong incentive.

It was found that only a very small number of children selected the peer approval or independence rewards. An analysis showed that the rewards actually offered, a certificate to be shown to a best friend or free play time (which all subjects received anyway) were not perceived as being equal in value to the other offered incentives (candy bar, certificate of merit, name on public bulletin board). No effective way of increasing the value of the least favored categories has been found and it was not thought desirable to decrease the value of the others.

In order to strengthen the effect of the RPI, based on the pilot studies at Penn State, the following modifications were tested in an additional pilot study: 1) the Peer Approval and Independence categories were eliminated; 2) additional pairings of examples of the remaining three categories were included for a new total of 24 pairings; and 3) intermittent statements of encouragement were added to the task. These encouraging statements have special reference to the particular reward assigned to the subject and are designed to remind him of the reward for which he is working.

Preliminary results using this modified method indicate that it results in a larger difference in the dependent variables between the most and least preferred reward groups than did the original procedure.

Text of the Inventory
(Administered by Computer
Assisted Instruction)

Suppose you worked hard on an assignment and you think that you have done a good job. Which one of the two things below would you most like to have happen?

Teacher gives you a gold star.
Teacher gives you a cookie.

You will use the light pen on the right side of this screen to answer. If you choose the gold star, press the light pen to the box besides that sentence. If you would like the cookie best, press it to the box in front of that sentence. Use the light pen now.

(Subject uses the light pen. Depending upon the response, the computer would respond:)

Very good, you chose the gold star (cookie).
Now try these.

(The screen is erased and the following pairs are successively presented.)

Pretend you have done well in your work.
Which of these would you like?
Point the light pen to it.

1. Teacher writes "100" on your paper.
   Be first to finish your work.

2. A candy bar.
   Teacher writes "perfect" on your paper.

3. Have your paper put on the bulletin board.
   An ice cream cone.

4. Teacher writes "100" on your paper.
   A package of bubble gum.

5. Teacher writes "perfect" on your paper.
   Be the only one who can answer a question.

6. A package of bubble gum.
   Be first to finish your work.

7. Teacher writes "A" on your paper.
   Be the only one that can answer a question.

8. Teacher writes "Excellent" on your paper.
   An ice cream cone.

9. A candy bar.
   Teacher writes "A" on your paper.

10. Have only your paper shown to the class.
    Teacher writes "100" on your paper.

11. Teacher writes "Excellent" on your paper.
    Be first to finish your work.

12. A candy bar.
    Be the only one that can answer a question.

13. Have only your paper shown to the class.
    Teacher writes "Perfect" on your paper.

14. Be first to finish your work.
    A soft drink.

15. An ice cream cone.
    Teacher writes "Perfect" on your paper.

16. Teacher writes "A" on your paper.
    Have your paper put on the bulletin board.

17. A soft drink.
    Teacher writes "100" on your paper.

18. Have only your paper shown to the class.
    An ice cream cone.

19. Have your paper put on the bulletin board.
    Teacher writes "Excellent" on your paper.

20. Teacher writes "A" on your paper.
    A package of bubble gum.

21. Have only your paper shown to the class.
    A candy bar.

22. A soft drink.
    Teacher writes "Excellent" on your paper.

23. A package of bubble gum.
    Be the only one that can answer a question.

24. Have your paper put on the bulletin board.
    A soft drink.

ERIC

APPENDIX C

Examples of the Rewards Administered

EXAMPLES OF THE REWARDS ADMINISTERED

COMPUTER ASSISTED INSTRUCTION LABORATORY

REPORT CARD

FOR

A. STUDENT

For your work at the
COMPUTER ASSISTED INSTRUCTION
LABORATORY

you have been given
a grade of

A

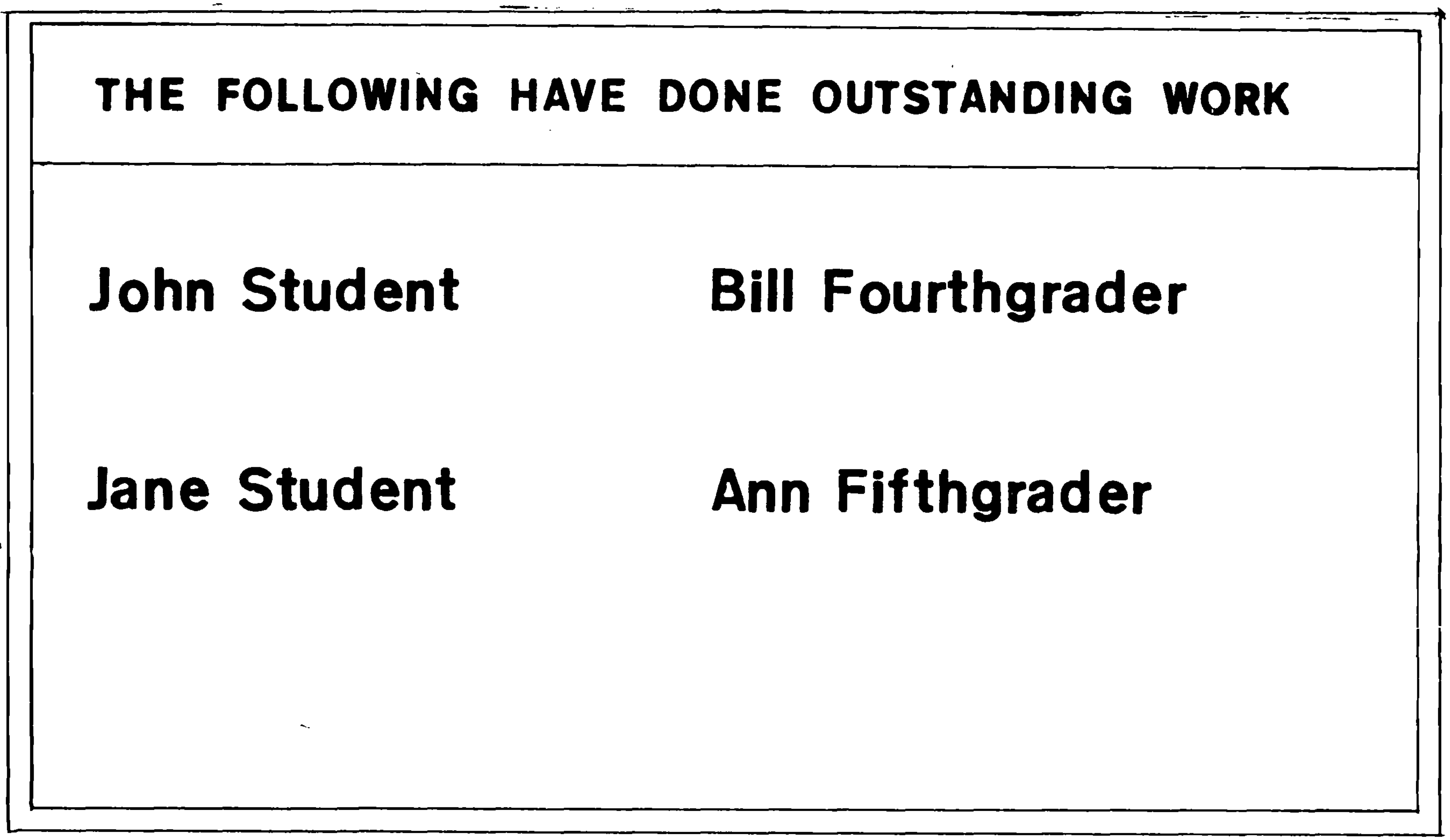
Meaning of grades:

A - Superior work
B - Good work
C - Average work
D - Fair work
F - Failing

THANK YOU FOR VISITING WITH US

David P. Yens

The Report Card
(Adult Approval)



The Chart on which Names were Posted
(Competitive Reward)

APPENDIX D

Description of Computer Assisted Instruction Facilities

ERIC

DESCRIPTION OF COMPUTER ASSISTED INSTRUCTION FACILITIES

The Computer Assisted Instruction Laboratory at Penn State has been in existence since 1964. Since that time the Laboratory has grown from a staff of four part-time faculty members to a present total of 65 University employees (faculty, graduate assistants, technicians, and clerical staff), equivalent to 46 full-time persons.

Quantity and sophistication of equipment has also grown considerably from a single teleprocessing typewriter terminal to a self-contained CAI system.

In December, 1967, the CAI Laboratory acquired and installed the first computerized system designed especially for individualized tutorial instruction--the IBM 1500 Instructional System. This system is located in rooms 201, 202 and 102 Chambers Building on the main Penn State campus at University Park, Pennsylvania. The system consists of 11 instructional stations with cathode-ray tube display, light pen, typewriter keyboard, image projector, and tape player/recorder. The computer room equipment, also in 201 Chambers Building, is comprised of an 1131 central processing unit, 1132 printer, 1442 card reader and punch, 1133 multiplexer control unit, two 2310 disk storage drives, 1502 station control, four 1518 typewriters, two 2415 tape drives, and 029 keypunch equipment.

The central processor is an IBM 1130 computer with 32,768 sixteen bit words of core storage. In addition to the usual peripheral equipment, the central processor depends upon five IBM 2311 disk drives (2,560,000 words) for the storage of usable course information and operating instructions. Twin magnetic tape drives record the interaction between the program and the student for later analysis and course revision. Core storage cycle time is 3.6 microseconds and read/write time for disk storage is 27.8 microseconds per word.

Each IBM 1500 student station consists of four optional display/response devices which may be used individually or in combination. The central instrument connected to the computer consists of a cathode-ray tube screen with sixteen horizontal rows and forty vertical columns for a total of 640 display positions. Information sufficient to fill the screen is available in microseconds from an internal random access disk. A light pen device enables the learner to respond to display letters, figures, and graphics by touching the appropriate place on the screen. A part of the CRT device is a typewriter-like keyboard which makes it possible for the learner to construct responses, have them displayed at any author-desired point on the CRT screen, and receive rapid feedback in the form of an evaluative message. Four dictionaries of 128 characters each of the course author's own design are capable of being used simultaneously. Thus, utilizing CAI, it would be technically feasible in a mathematics course to teach by means of four different sets of symbols simultaneously. An image projector loaded with a 16mm microfilm is capable of holding 1000 images on a single roll and of accessing forty images per second under program control. An audio play/record device based on four channels on a 1/4-inch tape permits rapid accessing of audio messages stored anywhere on the tape. This device is essential for language instruction and the play/record feature enables the student to compare and evaluate his production of language sounds with a model provided on the tape by the course author. An electric typewriter on the system is a separate device which enables the student to receive a hard copy of the interaction or dialogue between himself and the computer. The configuration of the CAI system is shown in Figure 12.

An additional, similar, IBM 1500 system has been installed in a van. This mobile facility can be moved to any desired location and be operational within a few hours. All equipment is located in the van and students work in the van at 14 stations equipped as described above. The use of this van permits the establishment of a CAI facility at any location for as long as desired and affords the opportunity to bring the CAI programs to the desired population rather than requiring the students to come to one location.

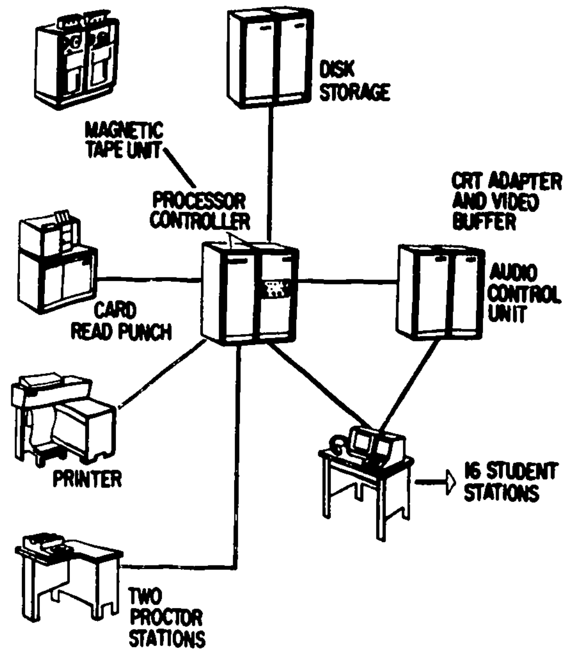


Fig. 12. Configuration of the CAI System

APPENDIX E

Content of the Pretest Administered by the Computer

ERIC

CONTENT OF THE PRETEST ADMINISTERED
BY THE COMPUTER

(display screen text)

Now I am going to find out how many kinds of arithmetic problems you can do.

I am going to give you some different kinds of problems to do. Some will be very easy and some will be very hard.

Some problems you may not have seen before. You probably will not be able to answer all of them.

Press the space bar to go on.

(display screen text, page 2)

You will have about a minute to do each problem, so work as fast as you can. Solve as many problems as you can.

If there are some answers you do not know, type an 'n' (find the 'n' on the keyboard) where the answer would go.

Type and enter the answers like you did with the TIC-TAC-TOE game. Look at the printed sheet for the way.

Press (the space bar)

(The display screen is erased and the problems are displayed one by one. Correct answers are followed by the word RIGHT centered on the lower part of the screen. Wrong answers are followed by WRONG, n responses by the word O.K. Each problem is printed in the format in which it was presented, and is accompanied by its difficulty level. Note that there are no levels 1 and 4. Overtimes are followed by SORRY, TIME'S UP. LET'S GO ON.)

(2) ADD

37 + 9 = _

(3) SUBTRACT

$$\begin{array}{r} 65 \\ \underline{-24} \end{array}$$

(5) ADD

$$\begin{array}{r} 7691 \\ \underline{+\ 5038} \end{array}$$

(6) DIVIDE

28 ÷ 4 = _

(7) SUBTRACT

$$\begin{array}{r} 342 \\ \underline{-263} \end{array}$$

(8) MULTIPLY

$$\begin{array}{r} 415 \\ \underline{\times\ 6} \end{array}$$

(9) SUBTRACT

$$\begin{array}{r} 8721 \\ \underline{-5893} \end{array}$$

(10) ADD

$$\frac{2}{5} + \frac{3}{5} = \square\frac{\square}{\square}$$

To see how to answer this question, look at page 3 of the booklet. Enter each part of the fraction at a time.

Raise your hand if you want help. (The booklet indicated the method of entering the response.)

(11) MULTIPLY

$$\begin{array}{r} 64 \\ \underline{\times 77} \end{array}$$

(12) SUBTRACT

$$\frac{4}{7} - \frac{2}{7} = \square\frac{\square}{\square}$$

(13) DIVIDE

$9\overline{)132}$ Remainder _

To answer this, type and enter the quotient first, then type and enter the remainder.

(14) MULTIPLY

$$\begin{array}{r} 479 \\ \underline{\times 68} \end{array}$$

(15) DIVIDE

$71\overline{)897}$ Remainder _

(16) SUBTRACT

$$\frac{7}{8} - \frac{3}{4} = \square\frac{\square}{\square}$$

(17) ADD

$$6\frac{1}{4} + 2\frac{3}{8} = \square\frac{\square}{\square}$$

(18) MULTIPLY

$$\frac{5}{8} \times \frac{3}{10} = \square\frac{\square}{\square}$$

(19) SUBTRACT

$$\frac{13}{17} - \frac{15}{34} = \square\frac{\square}{\square}$$

(20) MULTIPLY

$$2\frac{4}{9} \times 3\frac{2}{7} = \square\frac{\square}{\square}$$

(21) DIVIDE

$$\frac{3}{8} \div \frac{3}{5} = \square\frac{\square}{\square}$$

(22) DIVIDE

$$\frac{8}{3} \div \frac{1}{4} = \square\frac{\square}{\square}$$

(23) SUBTRACT

$$52\frac{3}{5} - 19\frac{1}{3} = \square\frac{\square}{\square}$$

Fine. That ends the quiz. Some of it was pretty hard, wasn't it? You should not have been able to answer all the questions. Now lets go on to the next part.

APPENDIX F

Samples of Items and Flowchart of Item Analysis

ERIC

SAMPLES OF ITEMS AND FLOWCHART OF ITEM ANALYSIS

Samples of Items

Samples of items at most of the difficulty levels can be found in Appendix E. Samples from the two levels omitted in Appendix E and from pretest items that were not representative of the task items are given below. The number in parenthesis is the level. Presentation could be vertical or horizontal (for non-fraction problems), determined randomly.

(1) ADD

$$\begin{array}{r} 7 \\ \underline{+3} \end{array}$$

(4) ADD

825 + 802 = ____

(16) ADD

$$\frac{3}{5} + \frac{7}{9} = \frac{\square}{\square}$$

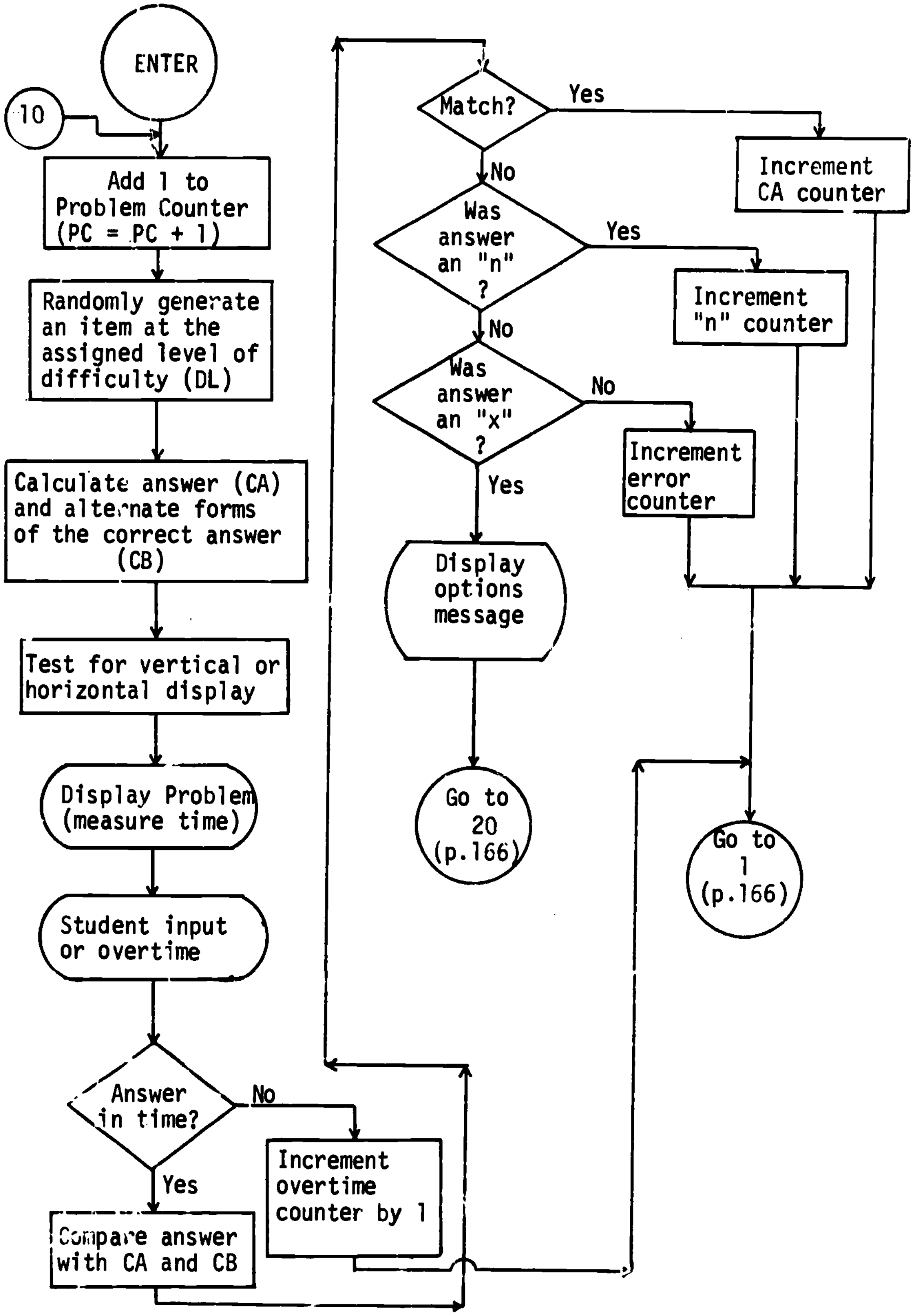
(22) DIVIDE

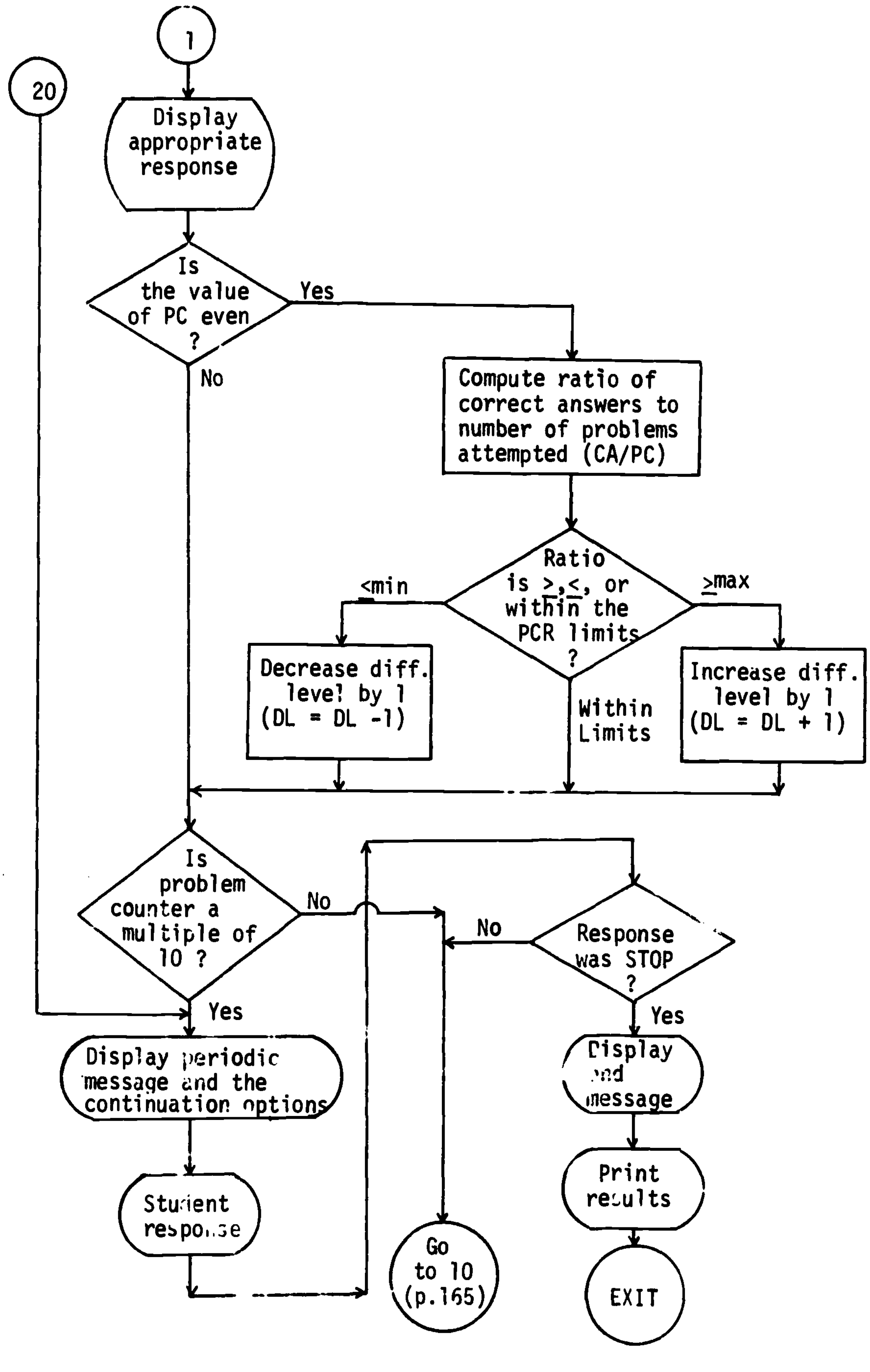
$$6\frac{2}{3} \div 1\frac{7}{8} = \square\frac{\square}{\square}$$

(23) SUBTRACT

$$4\frac{1}{6} - 2\frac{2}{5} = \square\frac{\square}{\square}$$

Flowchart of Item and Difficulty Level Processing





1
20
Display appropriate response
Is the value of PC even ?
Yes
No
Compute ratio of correct answers to number of problems attempted (CA/PC)
Ratio is >, <, or within the PCR limits ?
<min
>max
Decrease diff. level by 1 (DL = DL -1)
Within Limits
Increase diff. level by 1 (DL = DL + 1)
Is problem counter a multiple of 10 ?
No
Yes
Response was STOP ?
No
Yes
Display periodic message and the continuation options
Display end message
Student response
Print results
Go to 10 (p.165)
EXIT

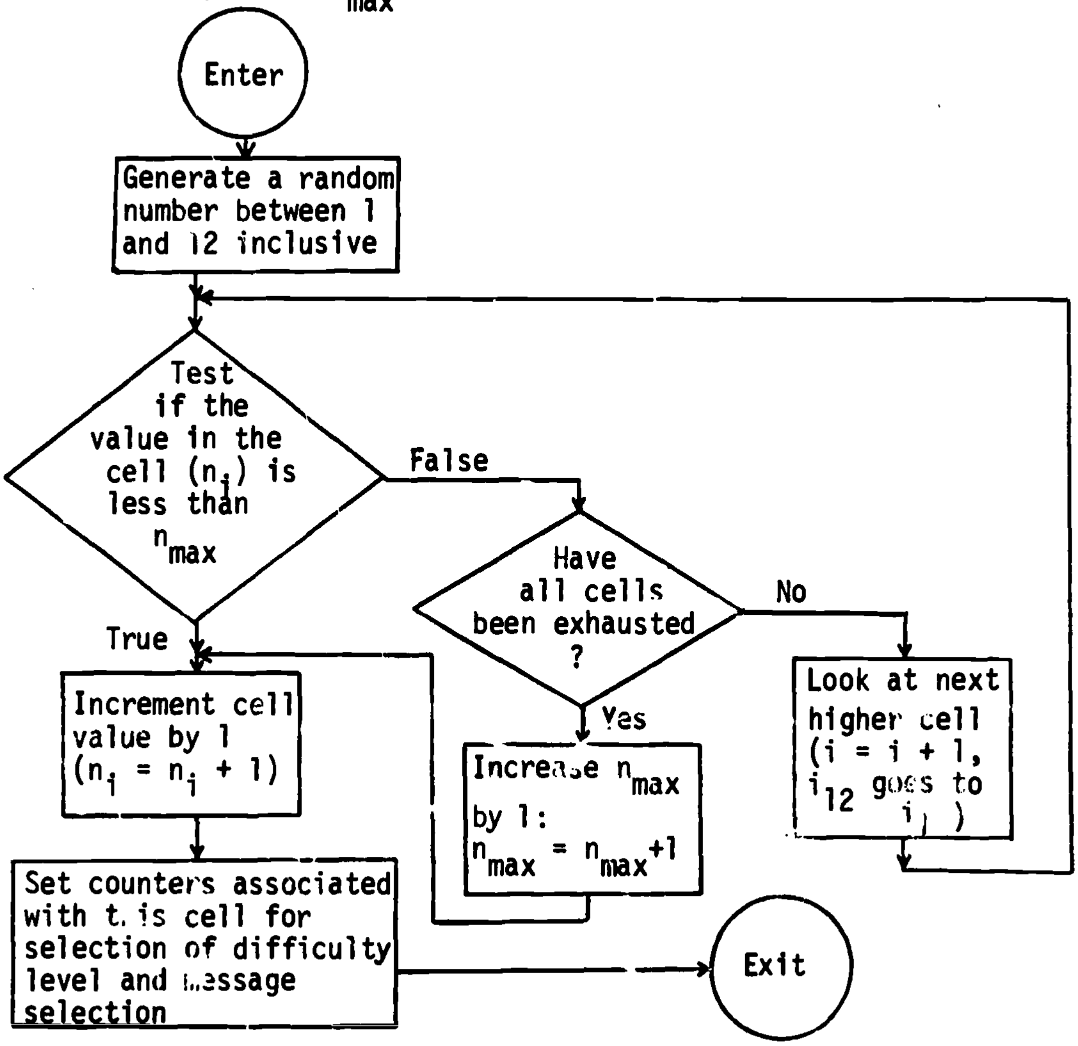
APPENDIX G

Methods for Assignment of Subjects to Cells and to the Initial Difficulty Level

METHODS FOR ASSIGNMENT OF SUBJECTS TO CELLS AND TO THE INITIAL DIFFICULTY LEVEL

Cell Assignment[3]

Assignment of each subject to one of the twelve cells was based upon maintaining a count of the number of subjects assigned to each cell (n_i, where i is the cell number) and the maximum number assigned to any cell (n_{max}).



[3]The process used was initially developed by Dr. Karl G. Borman of the Computer Assisted Instruction Laboratory, The Pennsylvania State University, and was modified by the author.

Assignment of Difficulty Level

The initial difficulty level (IDL) assigned to students depended upon the cell to which he was assigned (C1 = 98 - 80%, C2 = 72 - 54%, C3 = 46 - 28%, and C4 = 20 - 2% correct), the number of pretest problems solved correctly (PSUM), and the highest level of correctly answered pretest problem (PMAX).

If (PMAX - 2) is less than or equal to PSUM, then for

C1, IDL = PSUM - 3
C2, IDL = PSUM
C3, IDL = PSUM + 1
C4, IDL = PSUM + 3

If (PMAX - 2) is greater than PSUM, then for

C1, IDL = PSUM - 2
C2, IDL = PSUM
C3, IDL = PMAX + 1
C4, IDL = PMAX + 2

For example, if the number correct on the pretest was 5 and the highest level solved was 6, for a subject assigned to cell C3 the initial difficulty level would be set at 6 (5 + 1).

If PSUM was 8 and PMAX was 12, for cell C4 the IDL would be 14.

APPENDIX H

Typical Progress of a Selected Subject on the Task Problems

ERIC

TYPICAL PROGRESS OF A SELECTED SUBJECT ON THE TASK PROBLEMS

Subject Number 36 is a typical subject. This subject was assigned to cell 32 (No Reward, Moderately Easy Problems [PCR = 72 - 54%]). The specific complete data on the subject can be found on p. 184. The subject worked in the morning, starting the task problem at 9:34 and stopping at 10:37.

In the table below, the difficulty level, nature of response, and running percent correct (number correct/number attempted x 100) are given. Figure 13 is a graphical representation of the changes in difficulty level.

Table 19

Performance on Task Problems by a Typical Subject

Problem Number	Difficulty Level	Kind of Response	Percent
1	14	ca	100
2	14	wa	66
3	14	ca	75
4	14	ot	60
5	14	ca	66
6	14	ca	71
7	14	ca	75
8	15	n	66
9	15	n	60
	(message - continue)		
10	15	ca	63
11	15	n	58
12	15	n	53
13	15	ot	50
14	14	ca	53
15	14	wa	50
16	13	wa	47
17	13	n	44
18	12	ca	47

ERIC

Table 19 (Continued)

Problem Number	Difficulty Level	Kind of Response	Percent
19	12	ca	50
	(message - continue)		
20	11	ca	52
21	11	ca	54
22	10	wa	52
23	10	ca	54
24	9	ca	56
25	9	ca	57
26	9	ca	59
27	9	ca	60
28	9	ca	62
29	9	ca	63
	(message - continue)		
30	9	ca	64
31	9	ca	65
32	9	wa	63
33	9	ca	64
34	9	n	62
35	9	ca	63
36	9	ca	64
37	9	ca	65
38	9	ca	66
39	9	ca	67
	(message - continue)		
40	9	ca	68
41	9	wa	66
42	9	ca	67
43	9	ca	68
44	9	ca	68
45	9	ca	69
46	9	ca	70
47	9	ca	70
48	9	wa	69
49	9	ca	70
	(message - continue)		
50	9	ca	70
51	9	wa	69
52	9	ca	69
53	9	ca	70
54	9	ot	69
55	9	ca	69
56	9	ca	70
57	9	ca	70
58	9	ca	71
59	9	wa	70

Table 19 (Continued)

Problem Number	Difficulty Level	Kind of Response	Percent
	(message - continue)		
60	9	ca	70
61	9	ca	70
62	9	ca	71
63	9	ca	71
64	9	ca	72
65	9	ca	72
66	10	ca	73
67	10	ca	73
68	11	ca	73
69	11	ca	74
	(message - continue)		
70	12	ca	74
71	12	wa	73
72	13	ca	73
73	13	ot	72
74	14	ca	73
75	14	wa	72
76	15	n	71
77	15	ot	70
78	15	ca	70
79	15	n	70
	(message - continue)		
80	15	x	
	(message - stop)		

Meaning of Kind of Response Codes

ca = correct answer
wa = wrong answer
n = n entered (does not know how to solve)
ot = overtime; response not entered in time
x = stop option

Note that for the PCR range 72 - 54% a resulting net percent of 54 or less resulted in a decrease of difficulty level (easier problems) and a net percent of 72 or higher resulted in harder problems.

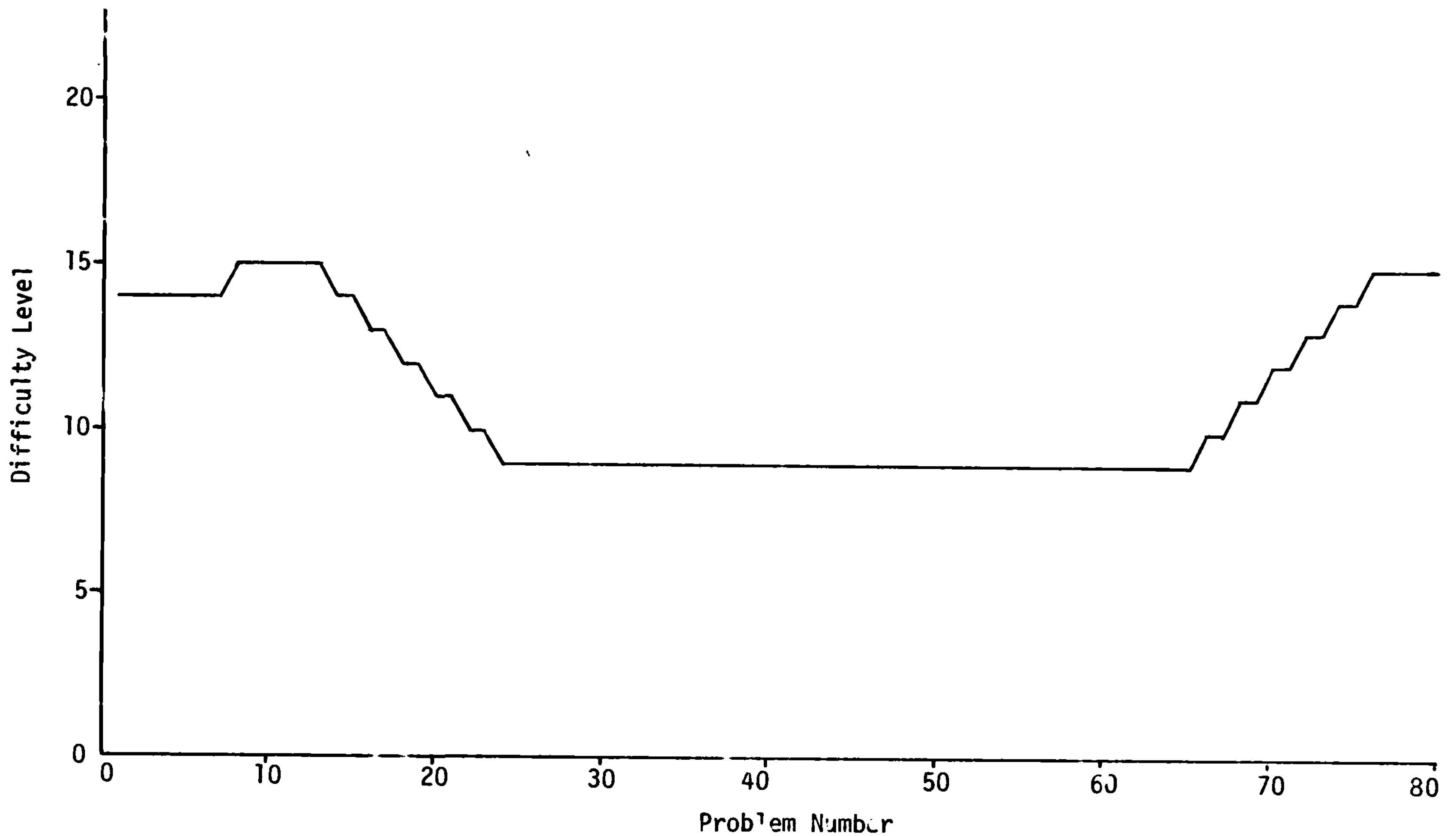


Fig. 13. Graph of Performance on the Task Problems by Subject Number 36 that Shows the Changes in Difficulty Level with Progress through the Problems

APPENDIX I

Data for All Subjects Used in the Analysis, by Cells

DATA FOR ALL SUBJECTS USED IN THE ANALYSIS, BY CELLS

Explanation of Variables and Levels on the Data Charts (From left to right)

Subj. No.: Assigned in chronological order

Rewd. Lvl.: Reward, or incentive, level (Independent Variable)

1 = Most Preferred

2 = Least Preferred

3 = None

Diff.: Difficulty, or Percent Correct Range (PCR, Independent Var.)

1 = Easy (98-80% correct)

2 = Moderately Easy (72-54%)

3 = Moderately Hard (46-28%)

4 = Hard (20-2%)

Sex:

1 = Male

2 = Female

Teach: Teacher

1 = Miss M. (5th grade)

2 = Mrs. S. (4th grade)

3 = Mrs. A. (5th grade)

4 = Mr. M. (4th grade)

5 = Mrs. B. (4th grade)

Grade:

1 = 4th

2 = 5th

TOD (Time of Day):

1 = a.m.

2 = p.m.

Rewd. Type (Kind of reward):

1 = Report card

2 = Name posted on chart

3 = M & M's

4 = None

Pref. (Preference)

Preference Value of the reward promised the subject - the number of times the category was selected during the Reward Preference Inventory.

Pretest:

ca: Number of pretest problems solved correctly.

Last: Difficulty level of the last pretest question solved correctly (the most difficult, since the test was constructed of problems in ascending order of difficulty).

Task and Performance Measures:

Start: Difficulty level of the first assigned problem.

Stop: Difficulty level of the last attempted problem.

Att: Number of problems attempted.

Ca: Number of problems answered correctly.

Wa: Number of wrong answer responses.

To: Number of overtimes (response time was restricted).

N: Number of problems answered with an "n" to indicate that S cannot do it.

Gross Percent: Overall percent of problems answered correctly.

Last Percent: Percent of the last ten problems answered correctly (a test of performance just prior to stopping).

TTime: Time spent working on the task only (Dependent Variable). (Time for reading reward messages and selecting continue/stop response is not included.) Recorded in seconds.

Latency: Summation of time spent making each response to a problem, measured from the time the problem is displayed until the entry of the response on the keyboard is completed (or the student goes overtime). Recorded in seconds.

Sec/Prob: Mean number of seconds required for each problem (TTime divided by number of problems attempted).

ATT-N: Number of problems attempted adjusted by subtracting the value for N (Dependent Variable).

Diff. Lvls. (Difficulty Levels):

#: Total number of changes in difficulty level during the task due to adjustment for maintenance of the prescribed error rate.

Range: The spread between and including the highest and lowest difficulty level encountered.

Force off:

Binary indication of whether the student terminated by himself (0) or whether he was asked to stop due to lunch or the end of school (1).

Validity: Arbitrary index of the validity of a subject's data.

1 = Minor discrepancy between a performance parameter and the actual performance; e.g., student was supposed to work in the 28-46% range but showed a 10% rate for the last 10 problems. Data acceptable.

2 = More serious discrepancy but data still usable. Example: a difference of more than 30% between assigned error rate and rate for last ten problems.

Data

Cell 11 (MPR - Easy Problems)

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff Lvls. #	Diff Lvls. Range	Force Off Validity
											Measured Task Variables														
59	11	1	2	1	1	3	14	2	12	1	4 1	09	89	20	0	0	82	70	28176	19274	177	109	4	4	
86	11	1	2	1	2	1	13	1	3	1	1 1	29	113	13	2	1	88	70	20352	12951	158	128	1	1	
74	11	1	3	2	1	1	14	2	5	2	2 1	35	122	12	1	0	90	83	36928	29561	274	135	1	1	1
129	11	1	4	1	2	1	12	3	6	3	1 1	69	159	8	2	0	94	90	25322	14617	150	169	3	3	
37	11	2	1	2	1	1	15	8	17	6	10 3	59	321	32	2	4	89	90	44538	18097	124	355	7	6	
021	11	2	1	2	1	2	10	06	13	04	01 1	48	130	18	00	00	88	090	33570	23947	227	148	4	4	
50	11	2	2	1	1	1	15	2	8	1	3 1	09	99	5		5	91	80	11305	6111	104	104	3	3	
118	11	2	4	1	1	1	16	4	12	2	4 1	70	147	17	6	0	87	91	62688	49659	369	170	3	3	
104	11	2	4	1	2	1	16	8	13	6	1 1	49	138	9	1	1	93	100	23423	15684	157	148	8	7	
132	11	2	5	1	1	1	15	9	13	7	4 1	14	102	7	5	0	90	93	45732	35879	401	114	8	6	
N=10		Means				-	14.	4.5	10.2	3.3	3.1	159.1	142.	14.1	1.9	1.1	89.	85.7	3320.3	2257.8	214.1	158.0	4.2	3.8	
																								Sum	1 0

Cell 21 (LPR - Easy Problems)

Measured Task Variables

Subj. No.	Rew Lvl. Diff.	Sex	Teach	Grade	TOD	Rewd. Type	Pref.	Pretest ca	Pretest Last	Start	Stop		Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff Lvls. #	Diff Lvls. Range	Force Off	Validity
93	21	1	2	1	1	2	6	1	2	1	1	1	69	138	31	0	0	82	40	17402	10045	103	169	3	2		1
92	21	1	2	1	2	2	6	3	5	2	1	1	57	150	4	2	1	96	100	28258	17046	180	156	8	5	1	
46	21	1	2	1	2	3	1	7	9	6	6	3	69	349	20	0	0	95	100	45276	23651	123	369	3	2	1	
120	21	1	4	1	1	2	4	3	12	1	1	1	39	134	5	0	0	96	100	22003	11804	158	139	3	2		
57	21	2	2	1	1	3	4	3	12	1	2	1	39	126	13	0	0	91	80	26067	18533	188	139	2	2		
31	21	2	2	1	2	3	3	7	9	6	1	1	68	147	20	0	1	88	90	23964	13019	143	167	6	6		
80	21	2	3	2	1	2	1	11	23	9	8	2	87	271	15	0	1	94	88	50202	38271	175	286	4	3	1	
51	21	2	3	2	1	3	4	5	9	3	1	4	79	445	29	4	1	93	100	57217	32707	119	478	3	3		
110	21	2	4	1	2	1	4	6	8	5	1		79	63	10	2	4	80	60	30531	24381	386	75	5	5	1	1
N=9		Means -					3.7	5.1	9.9	3.8	2.4		220.7	202.6	16.3	0.9	0.9	90.6	84.2	3343.6	2105.1	175.0	219.8	4.1	3.3		
																									Sum -	4	2

Cell 31 (No Reward - Easy Problems)

Measured Task Variables

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd. Type	Pref.	Pretest ca	Pretest Last	Start	Stop		Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff Lvls. #	Diff Lvls. Range	Force Off Validity
47	31	1	1	2	2	4		2	8	1	1	3	28	303	24	0	1	92	100	40140	29712	122	327	1	1	1
62	31	1	2	1	2	4		1	2	1	1	1	29	96	25	0	8	74	80	24639	10978	85	121	1	1	
103	31	1	4	1	2	4		3	12	1	1	3	19	310	9	0	0	97	100	32514	16612	102	319	1	1	
107	31	1	4	1	2	4		4	9	2	1	1	87	170	17	0	0	91	100	29940	17394	160	187	2	1	1
024	31	2	1	2	1	4	00	08	17	01	03	0	52	051	01	00	00	98	100	07296	04181	140	52	3	3	
41	31	2	1	2	2	4		7	16	5	11		59	51	4	2	2	88	80	24546	21216	416	57	7	7	
29	31	2	2	1	2	4		6	11	4	4		79	69	9	1	0	87	70	27738	22770	351	79	1	1	1
83	31	2	3	2	1	4		9	17	7	12	2	89	272	14	3	0	94	80	40030	22208	139	289	6	6	
139	31	2	5	1	1	4		10	14	8	8	2	09	175	8	10	16	84	100	48414	28882	232	193	15	8	
124	31	2	5	1	2	4		11	13	10	3	1	69	137	6	8	16	81	90	43510	29588	257	153	14	11	1
N=10		Means -						6.1	11.9	4.0	4.5		182.0	163.4	11.7	2.4	4.3	88.6	90.0	3187.7	2035.4	200.4	177.7	5.1	4.0	
																								Sum -		4 0

Cell 12 (MPR - Moderately Easy Problems)

Measured Task Variables

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	Gross N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
026	12	1	1	2	1	1	10	05	[illegible]	[illegible]	15	031	023	05	02	01	74	067	12684	09966	409	30	15	15		
64	12	1	2	1	2	1	14	2	3	3	9	89	61	17	3	1	69	40	32699	22317	251	88	11	9		
49	12	1	3	2	1	2	10	10	13	13	13	152	96	29	16	11	63	85	64811	52458	426	141	11	5		1
023	12	2	1	2	1	1	13	11	17	[illegible]	18	041	026	06	04	05	63	031	16712	13139	408	36	18	18	0	2
38	12	2	1	2	1	1	16	9	17	12	11	149	88	25	6	30	59	50	48024	36527	322	119	10	6		
019	12	2	1	2	2	1	13	10	13	13	17	041	027	03	08	03	66	070	21138	15517	516	38	5	5		1
87	12	2	2	1	2	1	13	2	6	5	7	9	5	4	0	0	56	56	3843	3329	427	9	3	3		
71	12	2	3	2	1	1	12	3	10	6	3	89	58	28	3	0	65	10	20638	16020	232	89	12	8		2
113	12	2	4	1	1	1	9	5	3	5	3	89	62	19	0	8	71	80	22824	17881	256	81	10	9		
126	12	2	4	1	2	1	10	2	3	3	3	142	96	46	0	0	68	69	42463	30799	299	142	7	2	1	
106	12	2	4	1	2	3	12	1	3	3	1	148	90	35	4	19	61	50	34680	24770	234	129	5	3	1	
136	12	2	5	1	1	1	13	2	10	5	5	119	75	33	3	8	63	50	48778	40780	410	111	5	3		
142	12	2	5	1	1	1	16	1	10	4	15	31	20	3	3	5	65	25	13325	9987	430	26	12	12		1
N=13		Means-12.4						4.8	9.4	[illegible].7	[illegible].2	[illegible]6.9	55.9	19.4	40	7.0	64.8	52.5	2943.2	2257.6	355.4	79.9	9.5	7.5		
																								Sum -	2	5

Cell 22 (LPR - Moderately Easy Problems)

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
025	22	1	1	2	1	1	03	12	17	01	18	036	024	06	06	00	67	020	14544	11487	404	36	18	18		
42	22	1	2	1	2	3	2	7	9	9	11	49	32	15	2	0	65	70	24234	20226	495	49	3	3		
73	22	1	3	2	1	3	1	4	9	7	11	164	102	14	3	45	62	27	35465	23235	216	119	9	5	1	1
114	22	1	4	1	1	3		9	13	12	10	55	37	11	2	5	67	40	19468	16204	354	50	9	6		1
130	22	1	4	1	2	2		1	2	2	4	29	18	10	1	0	62	50	8987	6635	310	29	5	4		
131	22	1	5	1	1	3		5	9	8	8	139	85	33	2	19	61	80	41735	27058	300	120	19	8		1
143	22	1	5	1	1	3	1	0	0	1	4	43	28	14	1	0	65	71	13321	9879	230	43	4	4		
101	22	1	5	1	1	3	2	4	12	7	7	33	18	8	7	0	55	36	15677	13535	475	33	5	3		1
007	22	2	1	2	1	3	01	07	13	10	13	047	032	05	00	10	68	089	15992	09139	340	37	12	4	1	2
018	22	2	1	2	2	3	04	06	17	09	04	049	034	07	01	07	69	070	18681	15189	381	42	8	7	1	
61	22	2	2	1	2	2	3	3	6	6	8	39	25	12	2	0	64	50	14124	10987	282	39	3	3		
33	22	2	2	1	2	3	3	0	0	1	10	20	10	6	2	2	50	55	10194	8504	510	18	10	10		
32	22	2	2	1	2	3	4	4	8	7	7	58	36	10	0	12	62	10	14844	10639	256	46	3	2		1
78	22	2	3	2	1	2	4	2	12	5	2	251	172	78	1	0	69	50	45670	28661	182	251	8	5	1	
N=14		Means-					2.0	4.6	9.1	6.1	8.4	72.3	46.6	16.4	2.1	7.1	63.3	51.3	2092.4	1509.8	338.2	65.1	8.3	5.9		
																							Sum -		4	6

Measured Task Variables

Cell 32 (No Reward - Moderately Easy Problems)

Measured Task Variables

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
36	32	1	1	2	1	4		11	16	14	15	79	55	11	5	8	70	40	35196	29590	446	71	14	7	1	1
34	32	1	2	1	2	4		1	6	4	2	48	28	11	0	9	58	30	9702	5060	202	39	5	4		1
67	32	1	3	2	1	4		12	18	10	1	29	22	3	2	2	76	70	9015	6781	311	27	12	12		
84	32	1	3	2	1	4		11	23	14	15	19	11	3	1	4	58	50	10304	8650	542	15	2	2		
90	32	1	3	2	1	4		3	12	6	7	159	112	21	8	18	70	60	55125	44753	347	141	6	3		
111	32	1	4	1	2	4		1	2	2	2	99	60	37	2	0	61	90	26208	21768	265	99	7	3	1	1
141	32	1	5	1	1	4		1	10	1	1	9	6	2	1	0	67	67	2615	1717	291	9	1	1		
147	32	1	5	1	1	4		5	10	8	6	59	38	7	1	13	64	70	9717	6741	165	46	5	3		
121	32	1	5	1	2	4		6	11	9	5	79	46	16	3	14	58	60	30819	21662	390	65	21	11		
022	32	2	1	2	1	4	00	06	17	09	15	039	028	11	00	00	72	055	11807	09064	303	39	15	11		
016	32	2	1	2	2	4	00	07	17	10	10	008	004	03	00	01	50	050	02057	01351	257	7	3	2		
60	32	2	2	1	2	4		2	8	5	2	29	18	11	0	0	62	50	6258	4675	161	29	6	5		
54	32	2	3	2	1	4		9	16	12	16	19	11	1	6	1	58	30	9312	7375	490	18	5	5		1
81	32	2	3	2	1	4		10	15	13	11	109	63	24	7	15	58	80	40301	32875	370	94	19	6		
95	32	2	4	1	1	4		6	13	9	6	109	72	12	22	3	66	100	49017	41269	450	106	24	11		1
N=15		Means-6.1						6.1	12.9	8.4	7.6	59.5	38.3	11.5	3.9	5.9	63.3	60.1	2049.7	1622.2	332.7	53 7	9.7	5.7		
																							Sum -		2	5

Cell 13 (MPR - Moderately Hard Problems)

Measured Task Variables

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff Lvls. #	Diff Lvls. Range	Force Off	Validity
014	13	1	1	2	2	3	13	01	02	03	02	018	005	12	00	01	28	10	07301	04780	406	17	4	2	0	1
43	13	1	2	1	2	1	14	2	8	9	7	29	10	10	1	8	35	20	8970	7149	309	21	5	4		
94	13	1	3	2	1	1	14	6	15	16	14	59	23	18	16	2	39	40	26766	22294	454	57	11	6		
77	13	1	3	2	1	1	15	10	16	17	14	50	14	8	8	20	28	18	19972	15638	399	30	10	5		1
127	13	1	4	1	2	1	15	2	5	6	5	99	29	34	6	30	29	30	28587	21581	289	69	20	10		
30	13	2	2	1	2	1	10	4	12	13	12	93	29	19	4	41	31	25	21468	13794	231	52	12	3		
45	13	2	2	1	2	1	13	4	8	9	12	61	15	13	3	30	25	25	17148	11308	281	31	22	13		
91	13	2	2	1	2	1	15	6	11	12	7	98	43	22	6	27	44	56	38808	31880	396	71	18	7	1	
72	13	2	3	2	1	3	10	6	12	13	13	96	37	18	2	39	39	10	17098	11617	178	57	7	2		1
117	13	2	4	1	1	1	16	5	10	11	15	139	44	13	46	36	32	0	65224	57147	469	103	13	4		1
134	13	2	5	1	1	1	12	8	12	13	11	178	83	48	3	44	47	78	47828	28291	269	134	31	10		1
N=11		Means				-	13.4	4.9	10.1	11.1	10.2	83.6	30.2	19.5	8.6	25.3	34.3	28.4	2719.7	2049.8	334.6	58.4	13.9	6.0		
																								Sum -	1	5

Cell 23 (LPR - Moderately Hard Problems)

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	Gross N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att-N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
															Measured Task Variables											
35	23	1	1	2	1	3	1	11	16	17	15	78	22	21	18	17	28	11	41550	34592	533	61	7	5		1
020	23	1	1	2	1	3	03	05	11	12	15	103	029	18	10	46	29	00	32877	25825	319	57	12	5	0	1
99	23	1	3	2	1	3		10	12	13	15	19	6	3	7	3	32	10	7765	6006	409	16	5	4		
105	23	1	4	1	2	1	7	4	13	14	14	129	46	8	9	66	36	0	33727	23379	261	63	9	3		1
125	23	1	4	1	2	3	1	2	9	10	3	79	27	28	3	21	34	10	18671	11806	236	58	12	10		1
66	23	2	2	1	2	3	2	6	12	13	14	51	18	8	4	21	35	0	14615	10470	287	30	6	3		1
109	23	2	4	1	2	3		5	12	13	5	121	47	44	10	20	39	0	37260	29296	308	101	17	13	1	1
145	23	2	5	1	1	3		8	13	14	17	179	57	36	6	80	32	0	46313	28567	259	99	20	6		1
N=8		Means -					1.8	6.4	12.3	13.3	12.3	94.9	31.5	20.8	8.4	34.3	33.1	3.9	2909.7	2024.3	326.5	60.6	11.0	6.1		
																								Sum -	1	7

Cell 33 (No Reward - Moderately Hard Problems)

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	#	Range	Force Off	Validity
								Measured Task Variables																		
40	33	1	1	2	2	4		6	12	13	15	9	3	1	0	5	33	33	2790	2116	310	4	3	3		
013	33	1	1	2	2	4	00	11	16	17	11	059	020	12	03	24	34	10	18680	13247	317	35	13	8	0	1
58	33	1	2	1	1	4		2	10	12	19	19	6	6	1	6	32	0	10201	5180	273	13	9	9		
88	33	1	2	1	2	4		2	6	7	10	18	6	7	1	4	33	33	6508	5510	362	14	4	4		
75	33	1	3	2	1	4		6	14	15	19	103	40	44	4	15	39	0	26367	17787	256	88	17	10	1	1
79	33	1	3	2	1	4		7	11	12	22	49	13	21	11	4	27	10	21455	17142	438	45	13	11		1
98	33	1	3	2	1	4		4	12	13	11	142	51	38	10	43	36	31	49006	38397	345	99	11	5		
112	33	1	4	1	1	4		10	13	14	20	29	8	17	0	4	28	10	9584	6785	330	25	7	6		1
123	33	1	4	1	2	4		5	8	9	4	69	29	32	7	1	42	30	30399	23873	441	68	8	7		
27	33	2	1	2	1	4		2	15	16	14	29	9	9	1	10	32	0	8658	6114	299	19	7	5		1
140	33	2	5	1	1	4		7	18	19	19	452	131	30	1	290	29	0	65022	24512	144	162	11	2		1
N=11		Means -						5.6	12.3	13.4	14.9	88.9	28.7	19.7	3.5	36.9	33.2	14.3	2260.6	1460.6	319.5	52.0	9.4	6.4		
																								Sum -	1	6

Cell 14 (MPR - Hard Problems)

Subj. No.	Rew. Lvl.	Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att-N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
													Measured Task Variables														
65	1	4	1	2	1	2	3	9	4	8	10	10	69	13	43	2	11	19	50	22111	14679	213	58	5	3		1
44	1	4	1	2	1	2	3	12	8	10	13	23	132	10	75	2	45	8	5	30078	19080	228	87	11	11		
28	1	4	2	2	1	2	3	10	10	12	15	15	60	1	8	11	40	2	0	20234	14954	337	20	1	1		
89	1	4	2	2	1	2	3	16	3	6	9	10	9	1	4	4	0	11	11	5048	4498	561	9	2	2		
70	1	4	2	3	2	1	1	11	7	15	17	17	85	4	65	8	8	5	0	29059	20407	342	77	1	1	1	
56	1	4	2	3	2	1	1	15	14	23	23	23	9	2	2	5	0	22	22	6433	4955	715	9	2	3		
116	1	4	2	4	1	1	1	9	5	9	11	18	179	32	94	29	24	18	80	61823	43991	345	155	10	9		1
128	1	4	2	4	1	2	1	11	4	6	9	9	87	13	23	19	32	15	6	41381	36294	476	55	1	1	1	
102	1	4	2	4	1	2	1	13	6	12	14	14	10	0	7	3	0	0	0	5347	4376	538	10	1	1		
144	1	4	2	5	1	1	1	14	9	11	14	14	105	11	19	12	63	11	13	45045	26914	429	42	7	4		
N=10							Means	-12.0	7.0	11.2	13.5	15.3	74.5	8.7	34.0	9.5	22.3	11.1	18.7	2665.6	1901.5	418.4	52.2	4.1	3.6		
																								Sum	-	2	2

Cell 24 (LPR - Hard Problems)

Measured Task Variables

Subj. No.	Rew. Lvl. Diff	Sex	Teach	Grade	TOD	Rewd Type	Pref.	Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
011	24	1	1	2	1	2	03	11	13	16	19	8	2	01	00	05	25	25	02369	01265	296	3	4	3		
008	24	1	1	2	1	3	01	05	17	18	13	24	1	08	10	05	04	09	14589	11316	608	19	12	6	1	1
63	24	1	2	1	2	2	6	1	2	5	15	109	4	28	8	69	4	0	27000	19913	183	40	11	11		
48	24	1	3	2	1	3		3	7	9	18	69	9	10	5	45	13	20	13966	11875	275	[illegible]	10	10		
76	24	1	3	2	1	3		6	9	12	20	89	3	30	2	54	3	0	18560	11146	210	35	9	9	1	
68	24	1	3	2	1	3	1	7	23	23	21	9	1	3	2	3	11	11	3510	2562	390	6	4	4		
82	24	1	3	2	1	3	2	8	13	15	15	19	0	4	15	0	0	0	15047	13382	704	19	1	1		
119	24	1	4	1	1	3	2	7	14	16	16	109	8	54	2	45	7	15	28986	15087	[illegible]	64	1	1		
39	24	2	1	2	1	3	0	8	16	18	19	129	9	114	6	0	7	0	33204	23695	257	129	2	2	1	
55	24	2	3	2	1	3	2	11	19	20	16	12	1	6	1	4	8	8	3924	2640	327	8	5	5		
97	24	2	4	1	1	3	1	4	8	10	23	179	16	98	4	61	9	0	47598	29013	266	118	14	14		
108	24	2	4	1	2	3	2	7	10	13	13	9	0	1	2	6	0	0	3999	3363	444	3	1	1		
137	24	2	5	1	1	3	1	8	13	15	13	135	25	17	1	92	19	13	25294	16286	187	43	3	3		
146	24	2	5	1	1	3	1	5	10	12	23	29	6	18	5	0	21	10	14090	12236	486	29	12	12		
N=14		Means					-1.6	6.5	12.4	14.4	17.4	66.4	6.1	28.0	4.5	27.8	9.4	7.9	1837.4	1241.3	349.9	38.6	6.4	5.9		
																								Sum -	3	1

Cell 34 (No Reward - Hard Problems)

Measured Task Variables

Subj. No.	Rew Lvl. Diff	Sex	Teach	Grade	TOD	Rewd. Type	Pref. Pretest ca	Pretest Last	Start	Stop	Att	ca	wa	To	N	Gross %	Last %	TTime	Latency	Sec/ Prob.	Att- N	Diff. Lvls. #	Diff. Lvls. Range	Force Off	Validity
85	34	1	2	1	2	4	2	3	6	11	39	2	12	1	24	5	0	8666	6289	222	15	6	6		
52	34	1	3	2	1	4	3	12	14	14	9	0	0	9	0	0	0	6689	5400	743	9	1	1		
53	34	1	3	2	1	4	4	18	16	19	9	1	5	2	1	11	11	3411	2125	379	8	4	4		
100	34	1	4	1	1	4	6	12	14	16	173	11	38	17	107	6	0	59120	41380	342	66	7	5		
9	34	2	1	2	1	4	5	13	15	15	21	0	2	7	12	0	0	10799	7646	514	9	1	1	1	
69	34	2	3	2	1	4	11	17	19	19	99	8	39	11	41	8	0	30542	21930	309	58	3	2	1	
115	34	2	4	1	1	4	9	12	15	14	109	8	38	49	14	7	0	66698	58922	612	95	2	2		
122	34	2	4	1	2	4	5	10	12	23	69	8	17	10	34	12	0	25832	16784	374	35	12	12		
133	34	2	5	1	1	4	8	13	15	14	129	15	39	8	67	12	80	49290	30148	382	62	12	7		
148	34	2	5	1	1	4	3	5	6	8	29	6	19	2	2	21	0	8878	7599	306	27	9	6		1
N=10		Means -					5.6	11.5	13.2	15.3	68.6	5.9	20.9	11.6	30.2	8.2	9.1	2699.3	1982.2	418.3	38.4	5.7	4.6		
																						Sum -		2	1

APPENDIX J

Rationale for Modification of Trend Analysis Computational Procedure

ERIC

RATIONALE FOR MODIFICATION OF TREND ANALYSIS COMPUTATION PROCEDURE

Because the curves of interest were the mean and the individual levels of incentive plotted over task difficulty, the method of orthogonal components for a one way design recommended by Winer (1962, p. 70-77) was selected for the tests of trend. However, this procedure failed to adequately account for the cell size variations found in the results. Because there were only four points in each curve and because of the use of column sums as the multiplier for the orthogonal coefficients, it was anticipated that the variations in cell sizes would unduly influence the results.

The harmonic mean is used for the calculation of the unweighted means analysis of variance. It was decided that the same assumptions that led to the selection of the unweighted means AOV (equal population sizes) would be applicable to the trend analysis. To implement this, the formulas provided by Winer (1962, p. 73) were modified to permit the use of the harmonic mean of marginal cell frequencies.

For <u>equal cell sizes</u>, the mean square for each component is computed using the variable C which is the sum of the products of the orthogonal coefficients and the column sums:

$$C = \Sigma\, T_i c_j$$

The variable D utilizes the number of observations for each column multiplier.

To adjust for unequal cell sizes, the following was done: adjustment for the column sums was made by using the product of the column means ($\bar{T}_j$) and the harmonic means of column cell frequencies ($\tilde{n}_h$) rather than the column sum. Thus, $C = \Sigma \tilde{n}_h \bar{T}_j c_j = \tilde{n}_h \Sigma \bar{T}_j c_j$.

For the calculation of D, the harmonic mean of the column cell frequencies replaced the simple number of obervations.

The error mean square used for computing the component F ratios was that for the unweighted means analysis of variance performed for the data subjected to trend analysis.

ERIC

APPENDIX K

Analysis of Variance Summary Tables and
Graphs of the Supplementary Analyses

ERIC

Table 20

Analysis of Variance Summary Table for Incentive, PCR, and Sex, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	2795970.	2	1397985.	<1.0	0.606
2 PCR	13472205.	3	4490735.	1.618	0.189
3 Sex	8024315.	1	8024315.	2.891	0.092
12	9287075.	6	1547846.	<1.0	0.763
13	502287.	2	251143.	<1.0	0.914
23	10701226.	3	3567075.	1.285	0.283
123	5391282.	6	898547.	<1.0	0.923
Error	308073522.	111	2775437.		

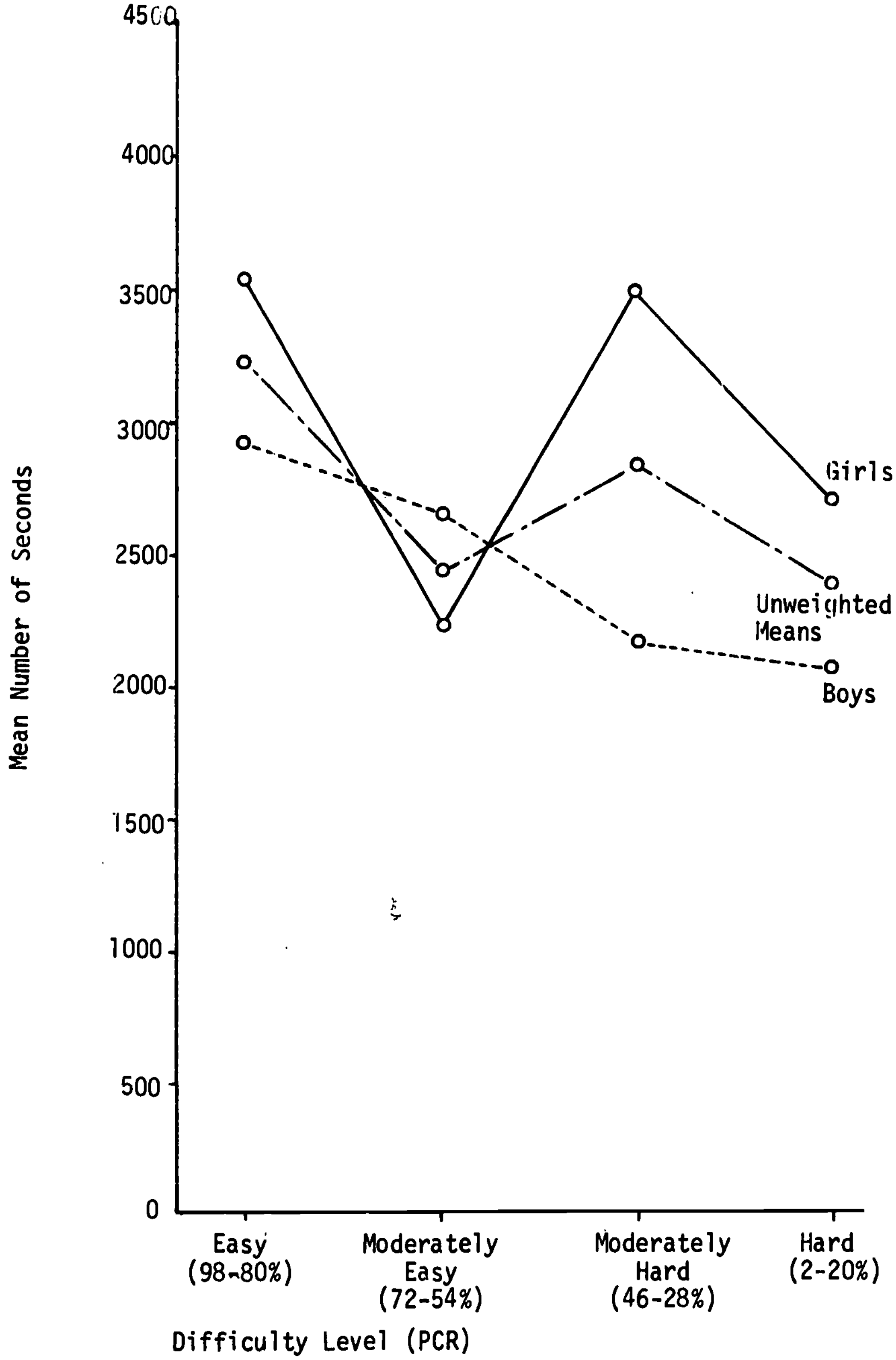


Fig. 14. Interaction between Sex and Task Difficulty (PCR) with Task Time as the Dependent Variable, One Aspect of the Three Way Interaction between Sex, Incentive, and Task Difficulty.

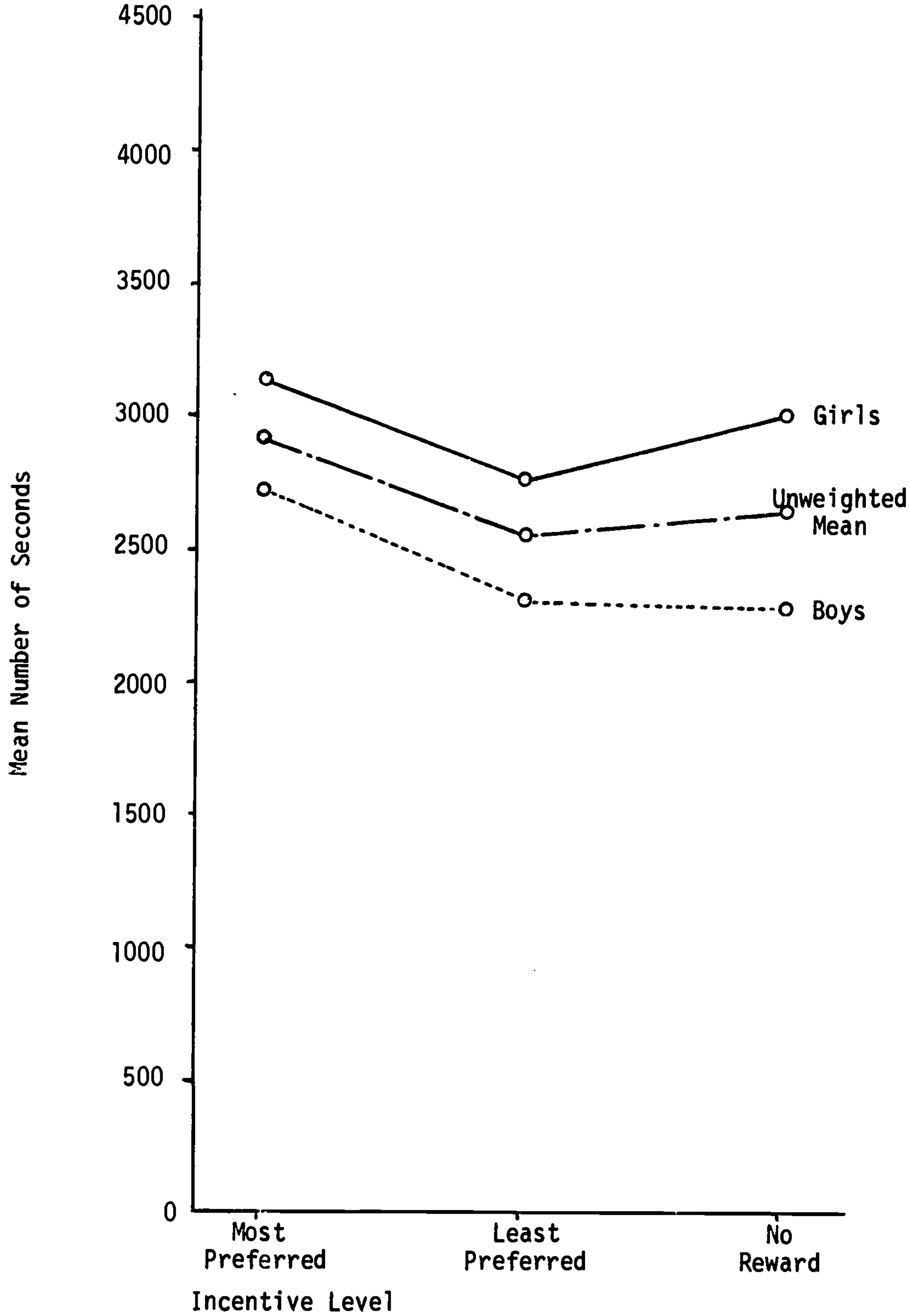


Fig. 15. Interaction between Sex and Incentive with Task Time as the Dependent Variable, One Aspect of the Three Way Interaction between Sex, Incentive, and Task Difficulty.

Table 21

Analysis of Variance Summary Table for Incentive, PCR, and Sex, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	2690.	2	1345.	<1.0	0.707
2 PCR	355048.	3	118349.	30.588	0.000
3 Sex	1522.	1	1522.	<1.0	0.532
12	24248.	6	4041.	1.044	0.401
13	4787.	2	2393.	<1.0	0.541
23	9653.	3	3218.	<1.0	0.479
123	27516.	6	4586.	1.185	0.319
Error	429481.	111	3869.		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 62.99 with 23 d.f., $p<.001$. The Hypothesis of Homogeneity of Variance is rejected.

Assuming proportional population subclass frequencies, the probability of obtaining a sample with the obtained subclass frequencies is less than 0.1.

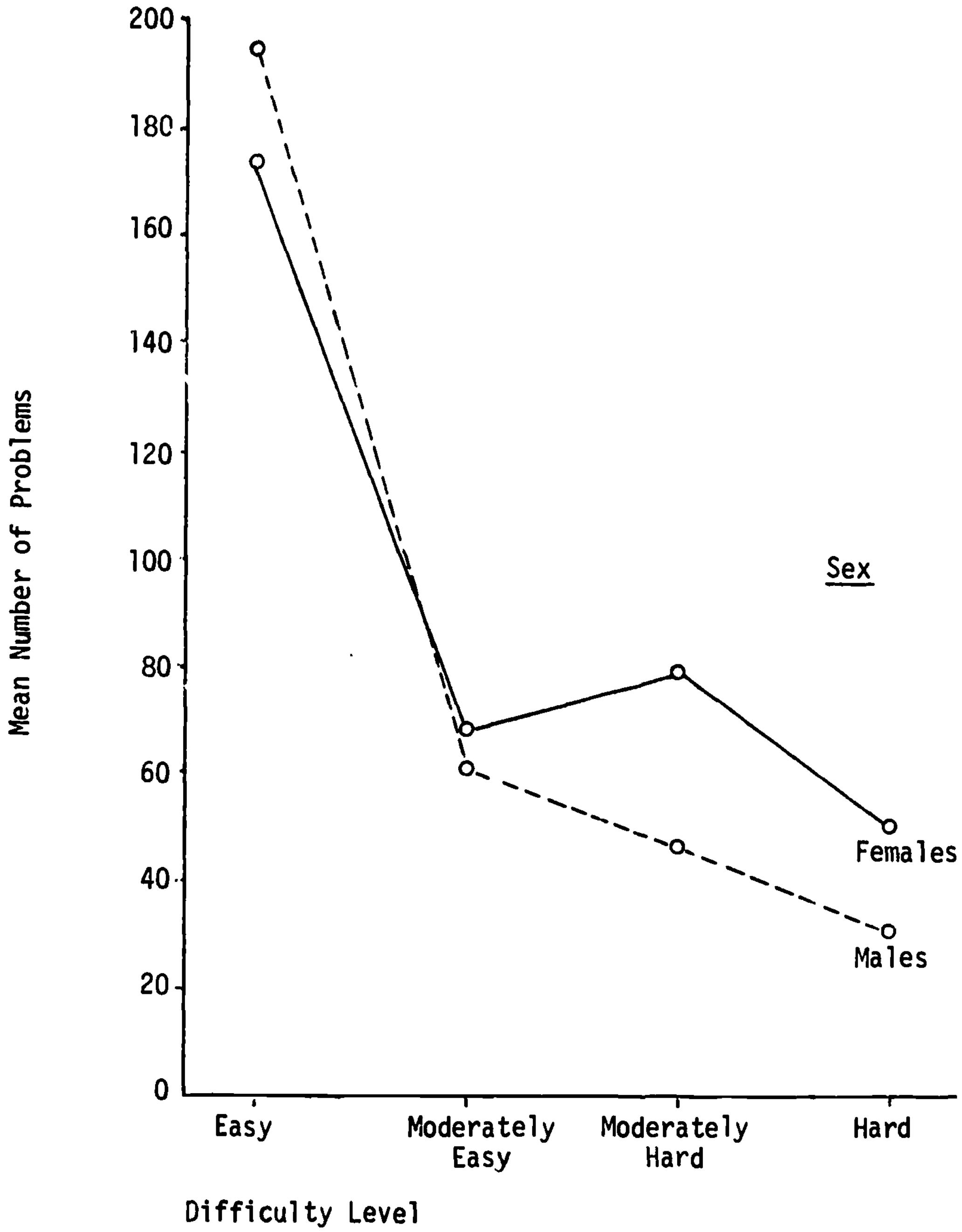


Fig. 16. Interaction between Sex and Task Difficulty (PCR) with ANP as the Dependent Variable, One Aspect of the Three Way Interaction between Sex, Incentive, and Task Difficulty.

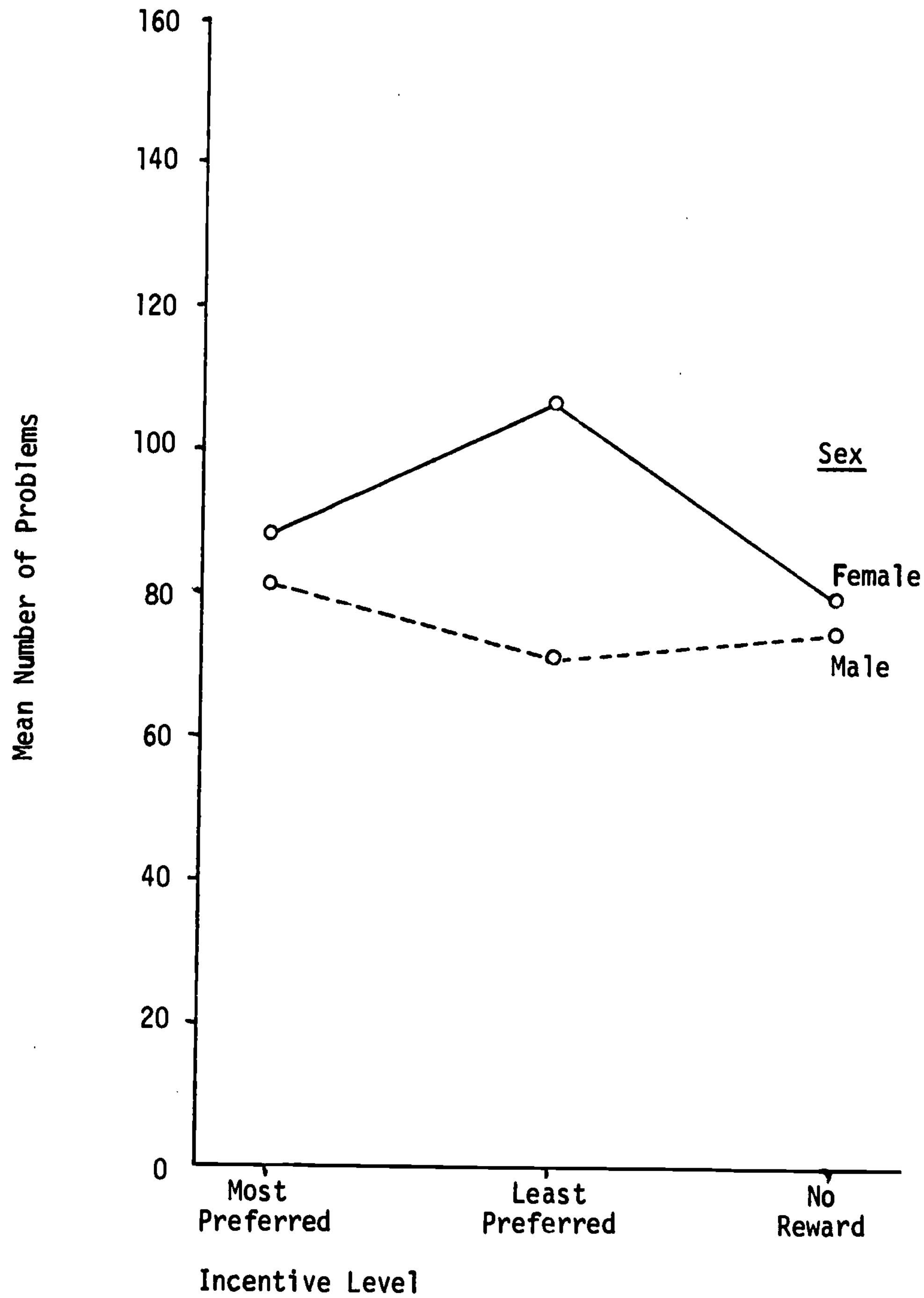


Fig. 17. Interaction between Sex and Incentive with ANP as the Dependent Variable, One Aspect of the Three Way Interaction between Sex, Incentive, and Task Difficulty.

Table 22

Analysis of Variance Summary Table for Incentive, PCR, and Grade, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	2576399.	2	1288200.	<1.0	0.614
2 PCR	29990659.	3	9996886.	3.808	0.012
3 Grade	1379160.	1	1379160.	<1.0	0.470
12	9540408.	6	1590068.	<1.0	0.725
13	8342702.	2	4171351.	1.589	0.209
23	25381551.	3	8460517.	3.223	0.025
123	9574316.	6	1595719.	<1.0	0.724
Error	291424622.	111	2625447.		

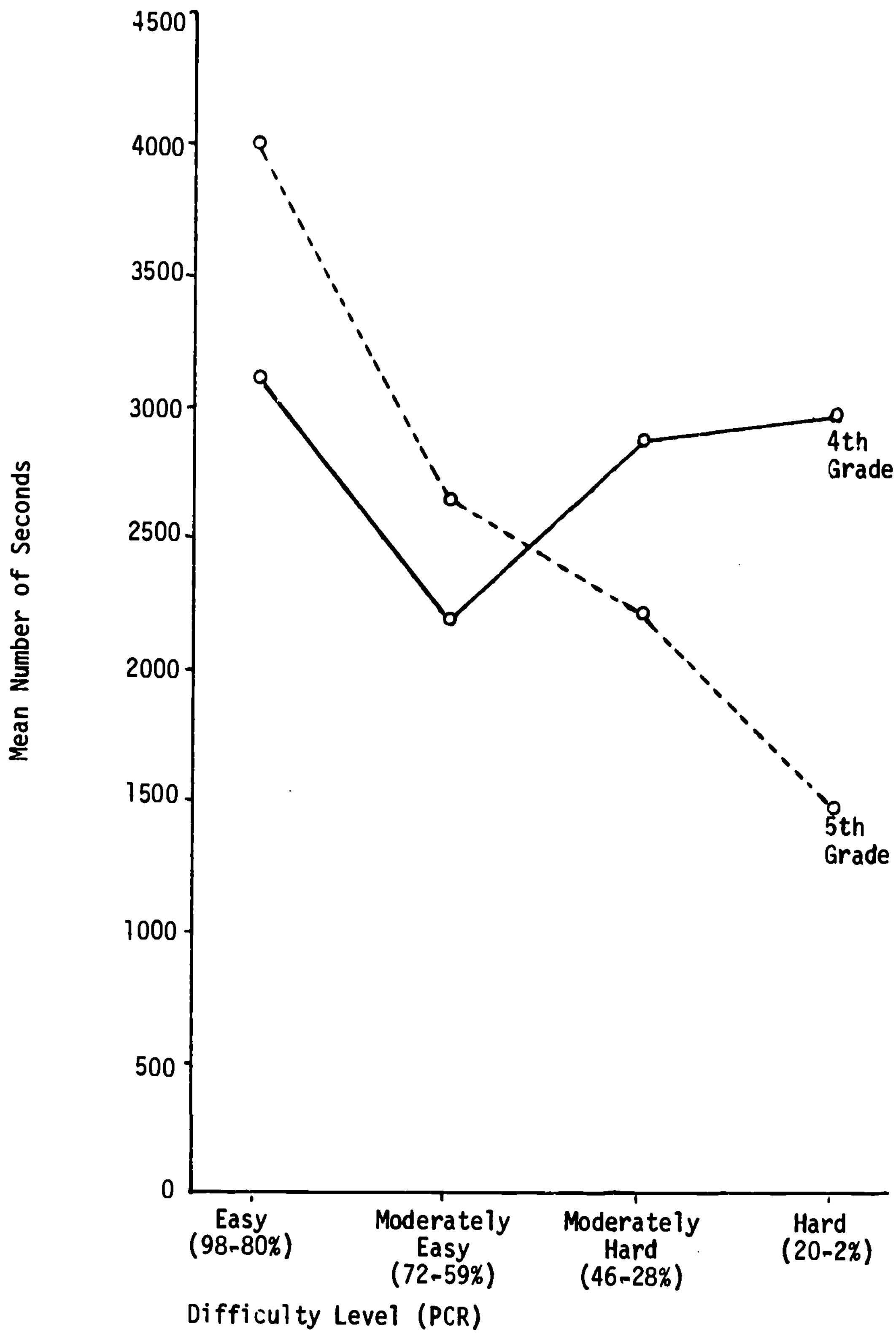


Fig. 18. Interaction between Grade and Task Difficulty (PCR) with Task Time as the Dependent Variable, One Aspect of the Three Way Interaction between Grade, Incentive, and Task Difficulty.

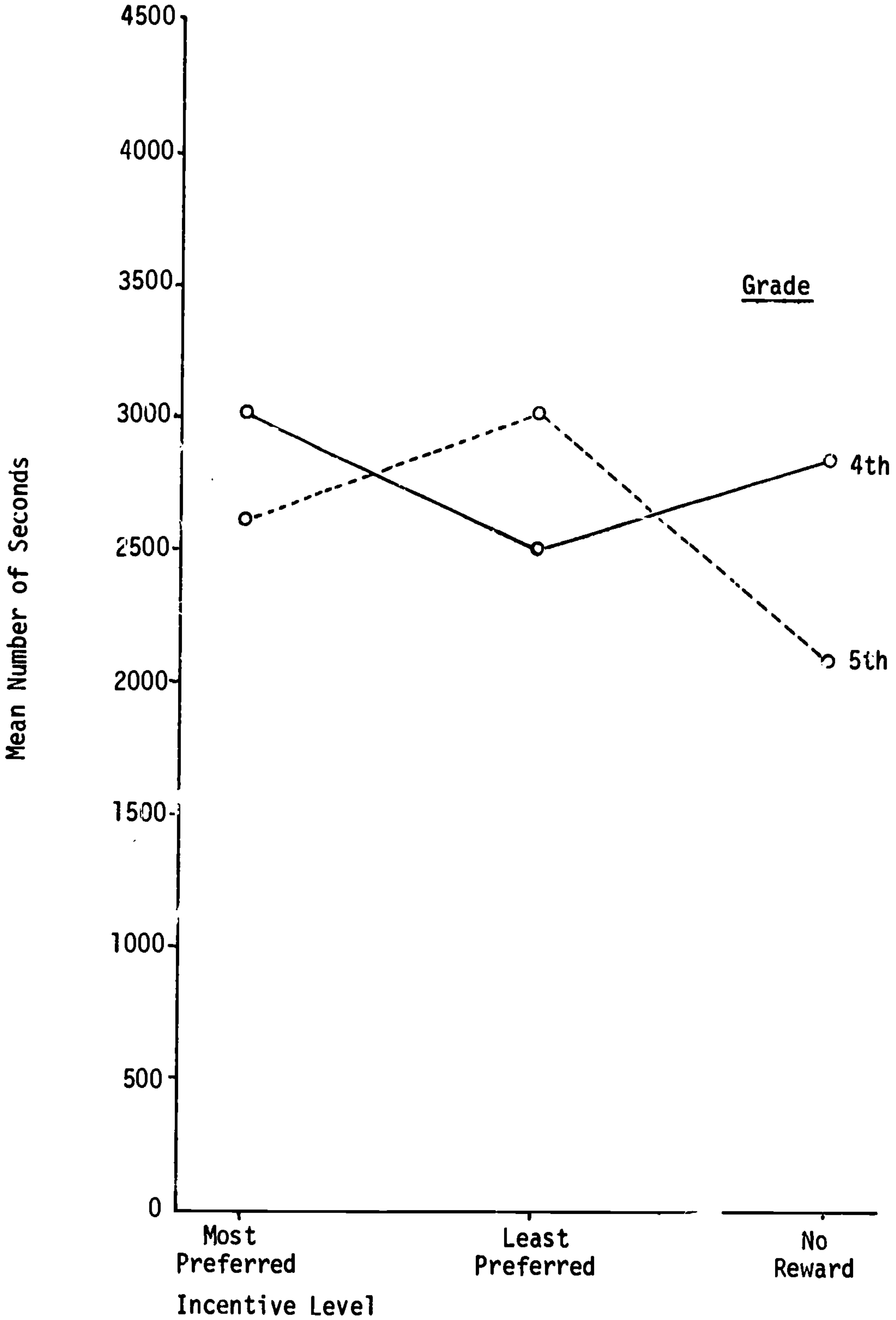


Fig. 19. Interaction between Grade and Ir entive with Task Time as the Dependent Variable, One Aspect of the Three Way Interaction between Grade, Incentive, and Task Difficulty.

Table 23

Analysis of Variance Summary Table for Incentive, PCR, and Grade, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	20431.	2	10216.	3.006	0.054
2 PCR	524094.	3	174698.	51.400	0.000
3 Grade	8533.	1	8533.	2.511	0.116
12	49622.	6	8270.	2.433	0.030
13	19828.	2	9914.	2.917	0.058
23	66201.	3	22067.	6.493	0.000
123	36232.	6	6039.	1.777	0.110
Error	377264.	111	3399.		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 51.52 with 23 d.f., $p<.001$. The Hypothesis of Homogeneity of Variance is rejected.

Assuming proportional population subclass frequencies, the probability of obtaining a sample with the obtained subclass frequencies is less than 0.5.

ERIC

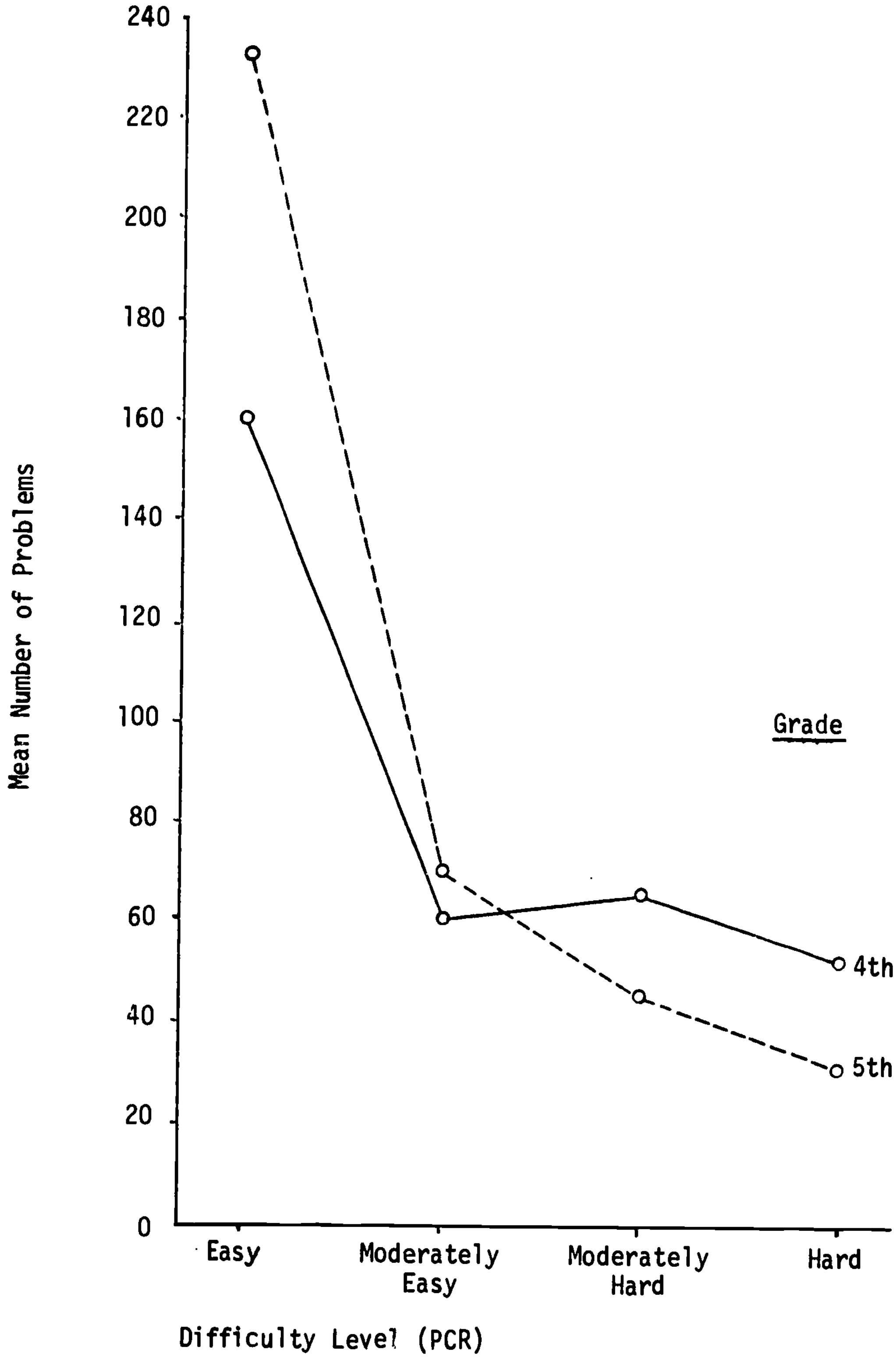


Fig. 20. Interaction between Grade and Task Difficulty (PCR) with ANP as the Dependent Variable, One Aspect of the Three Way Interaction between Grade, Incentive, and Task Difficulty.

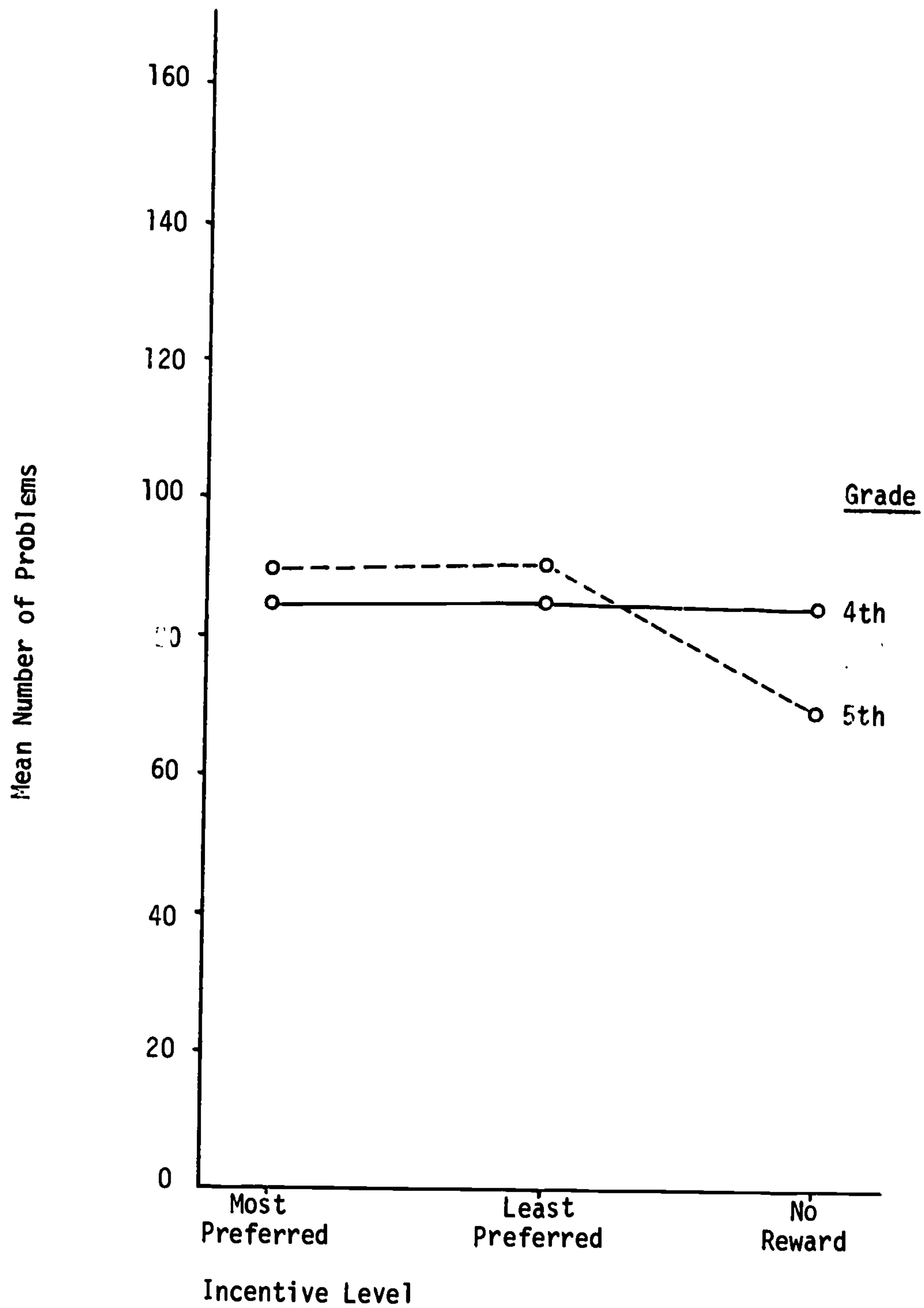


Fig. 21. Interaction between Grade and Incentive with ANP as the Dependent Variable, One Aspect of the Three Way Interaction between Grade, Incentive, and Task Difficulty.

Table 24

Analysis of Variance Summary Table for Incentive and PCR for Fifth Grade Students with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	6552333.	2	3276167.	1.422	0.252
2 PCR	37907147.	3	12635716.	5.483	0.003
12 Incentive x PCR	9813431.	6	1635572.	<1.0	0.644
Error	99091888.	43	2304463.		

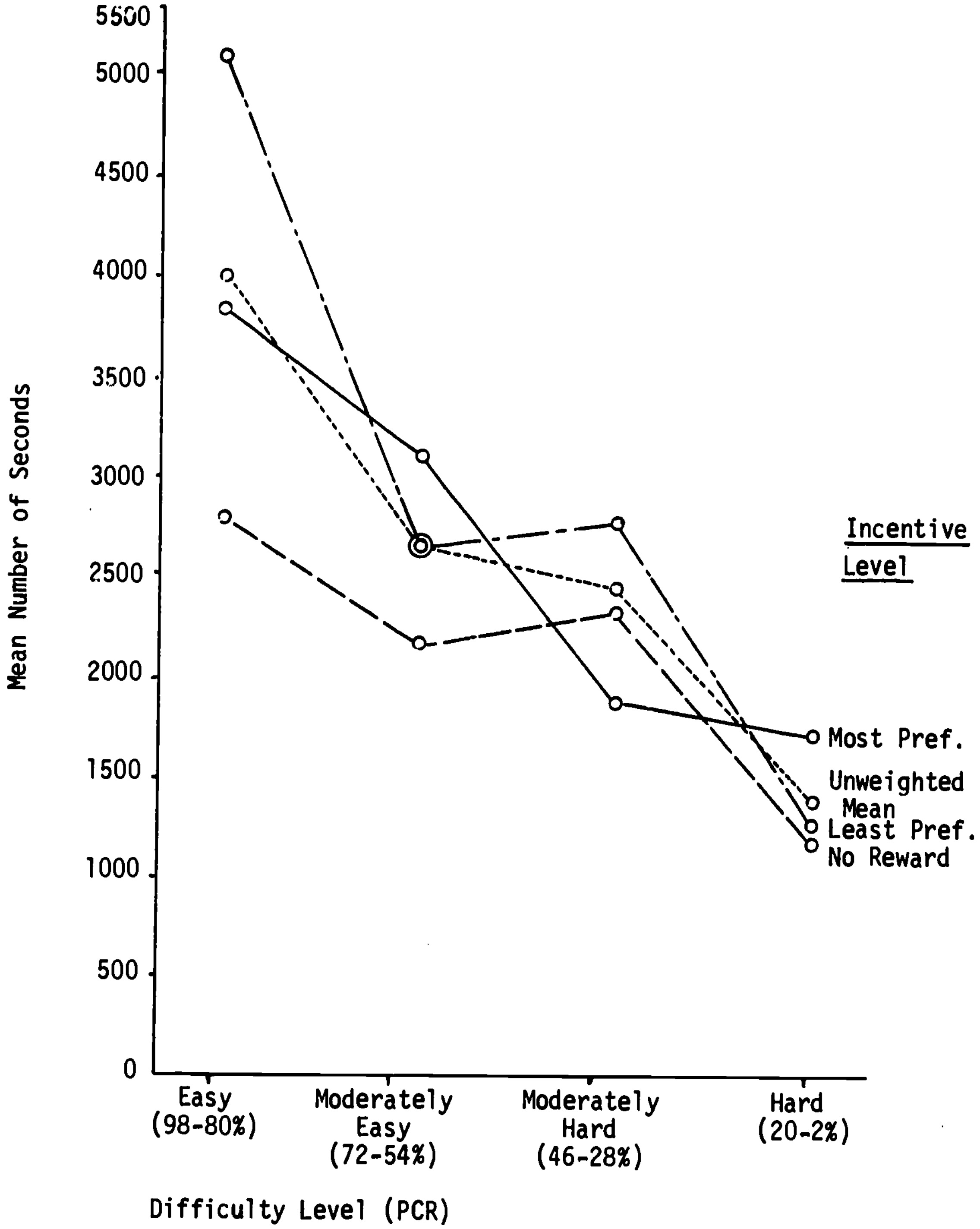


Fig. 22. Interaction between Incentive and Task Difficulty for Fifth Grade Data with Task Time as the Dependent Variable.

Table 25

Analysis of Variance Summary Table for Incentive and PCR for Fifth Grade Students with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	31763.	2	15882.	3.400	0.043
2 PCR	377713.	3	125904.	26.952	0.000
12 Incentive x PCR	60611.	6	10102.	2.162	0.065
Error	200870.	43	4671.		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 25.48 with 11 d.f., $p<.008$. The Hypothesis of Homogeneity of Variance is rejected.

ERIC

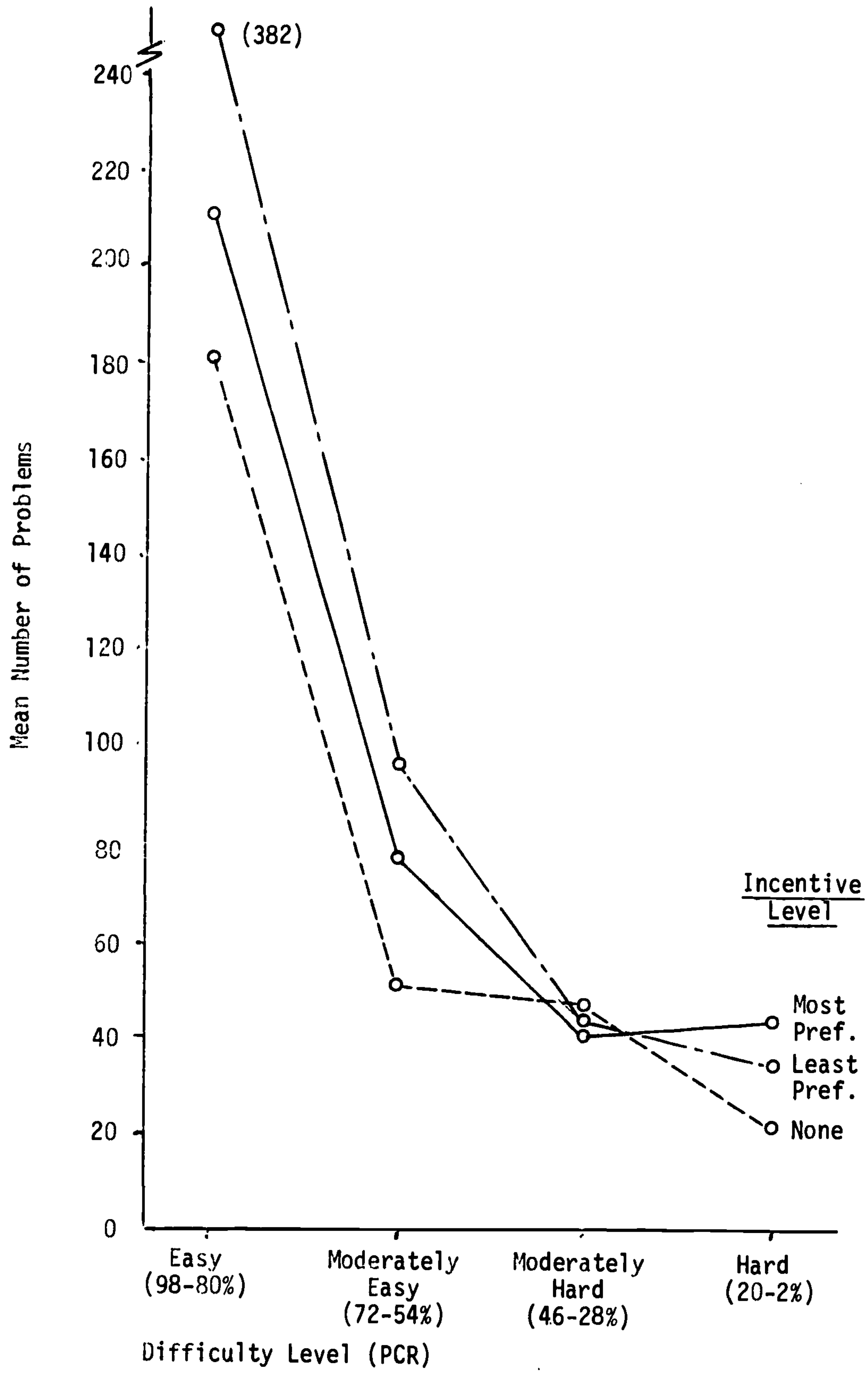


Fig. 23. Interaction between Incentive and Task Difficulty for Fifth Grade Data with ANP as the Dependent Variable.

Table 26

Analysis of Variance Summary Table for Incentive and PCR for Fourth Grade Students with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	3569738.	2	1784869.	<1.0	0.535
2 PCR	10010254.	3	3336751.	1.180	0.324
12 Incentive x PCR	9114526.	6	1519088.	<1.0	0.778
Error	192332734.	68	2828423.		

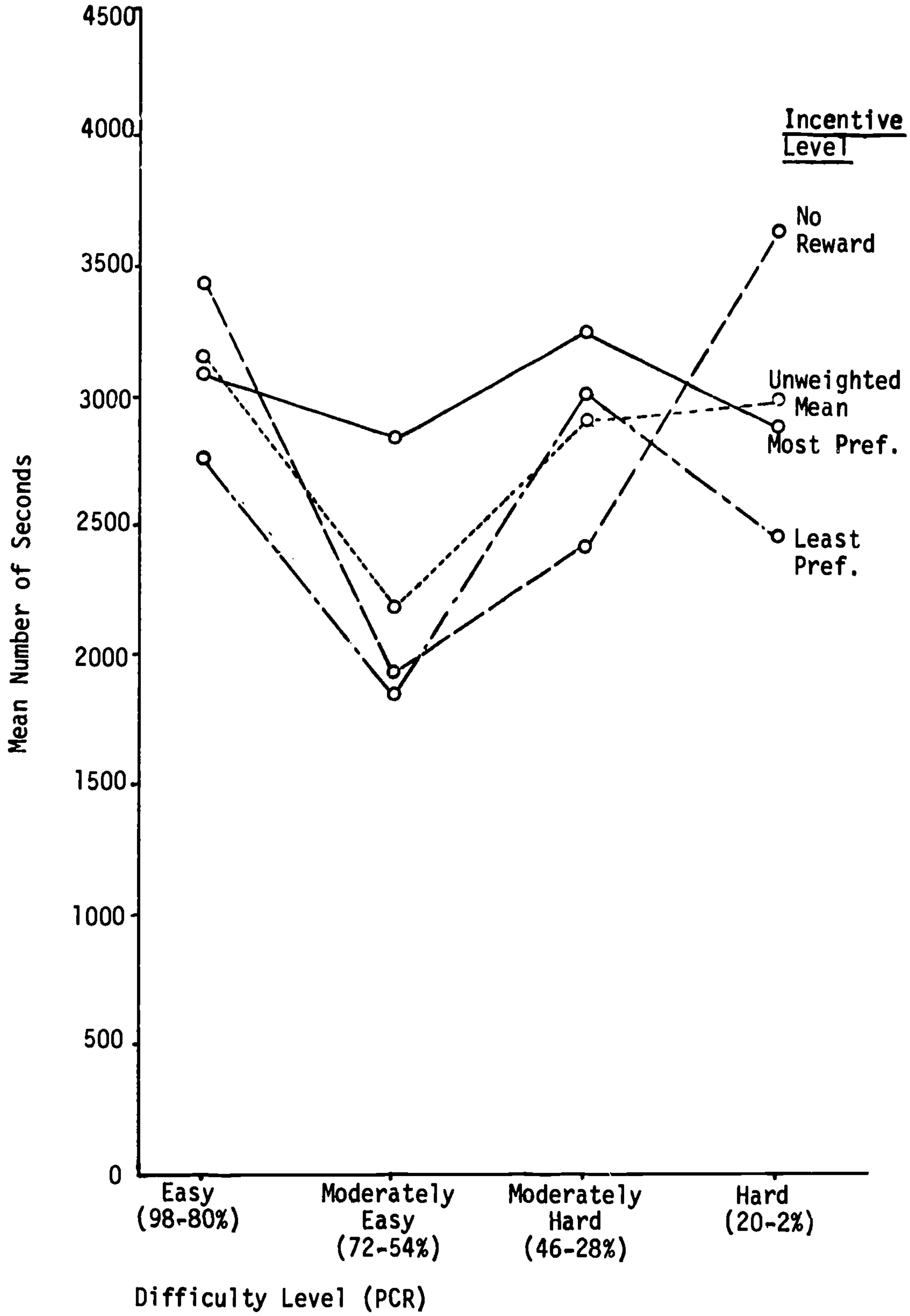


Fig. 24. Interaction of Incentive with Task Difficulty (PCR) for Fourth Grade Data with Task Time as the Dependent Variable.

Table 27

Analysis of Variance Summary Table for Incentive and PCR for Fourth Grade Students with ANP as the Dependent Variable

	Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1	Incentive	11.49	2	5.746	<1.0	0.998
2	PCR	152362.08	3	50787.360	19.579	0.000
12	Incentive x PCR	12345.16	6	2057.526	<1.0	0.578
	Error	176394.20	68	2594.032		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 21.45 with 11 d.f., $p<.029$. The Hypothesis of Homogeneity of Variance is rejected.

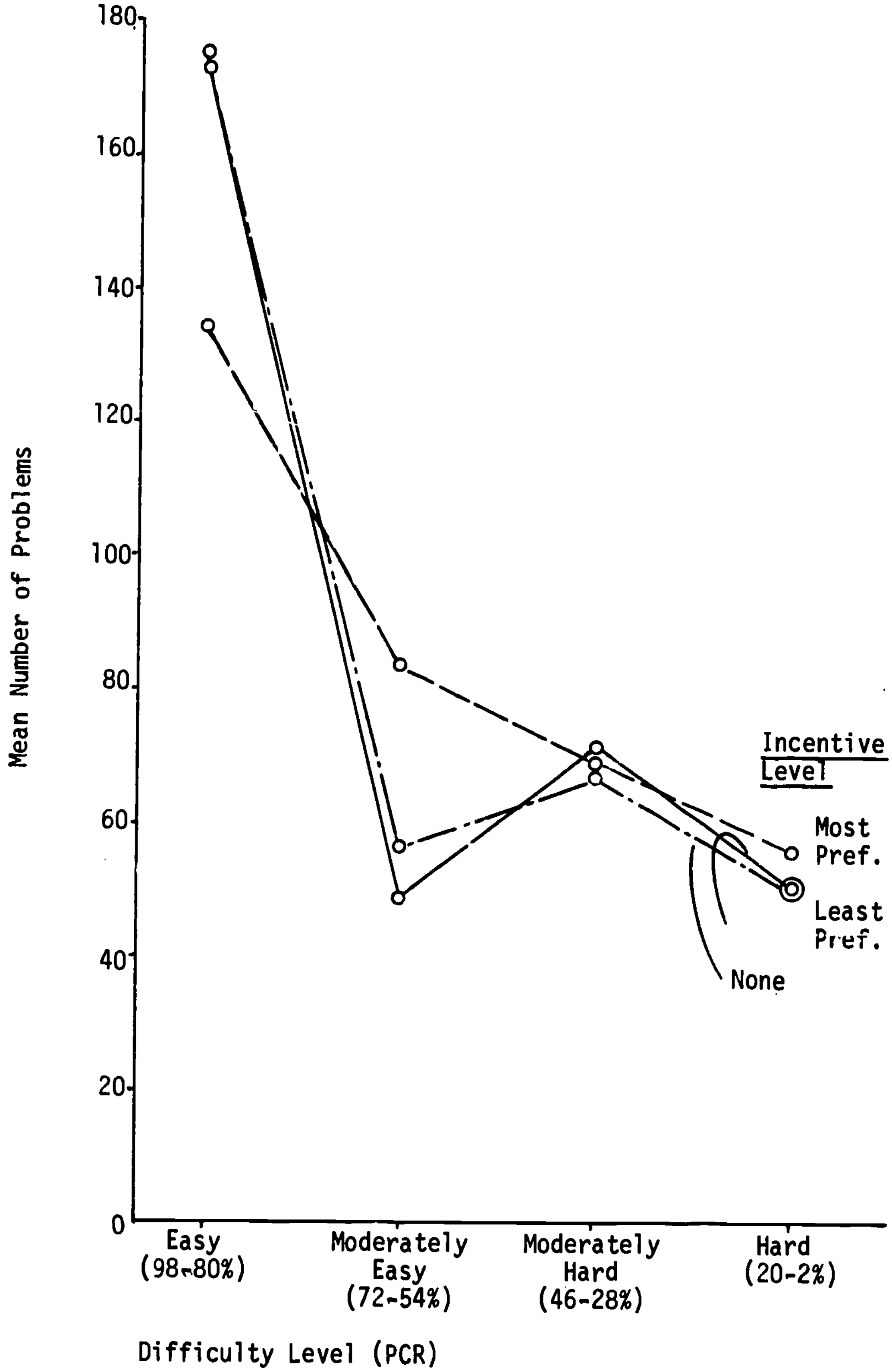


Fig. 25. Interaction of Incentive with Task Difficulty (PCR) for Fourth Grade Data with ANP as the Dependent Variable.

Table 28

Analysis of Variance Summary Table for Incentive, PCR, and Time of Day, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	5099632.	2	2549816.	<1.0	0.395
2 PCR	14946213.	3	4982071.	1.830	0.146
3 Time of Day	20594924.	1	20594924.	7.564	0.007
12	9888918.	6	1648153.	<1.0	0.726
13	2025445.	2	1012722.	<1.0	0.690
23	1306193.	3	435398.	<1.0	0.923
123	3842926.	6	640488.	<1.0	0.964
Error	302206226.	111	2722579.		

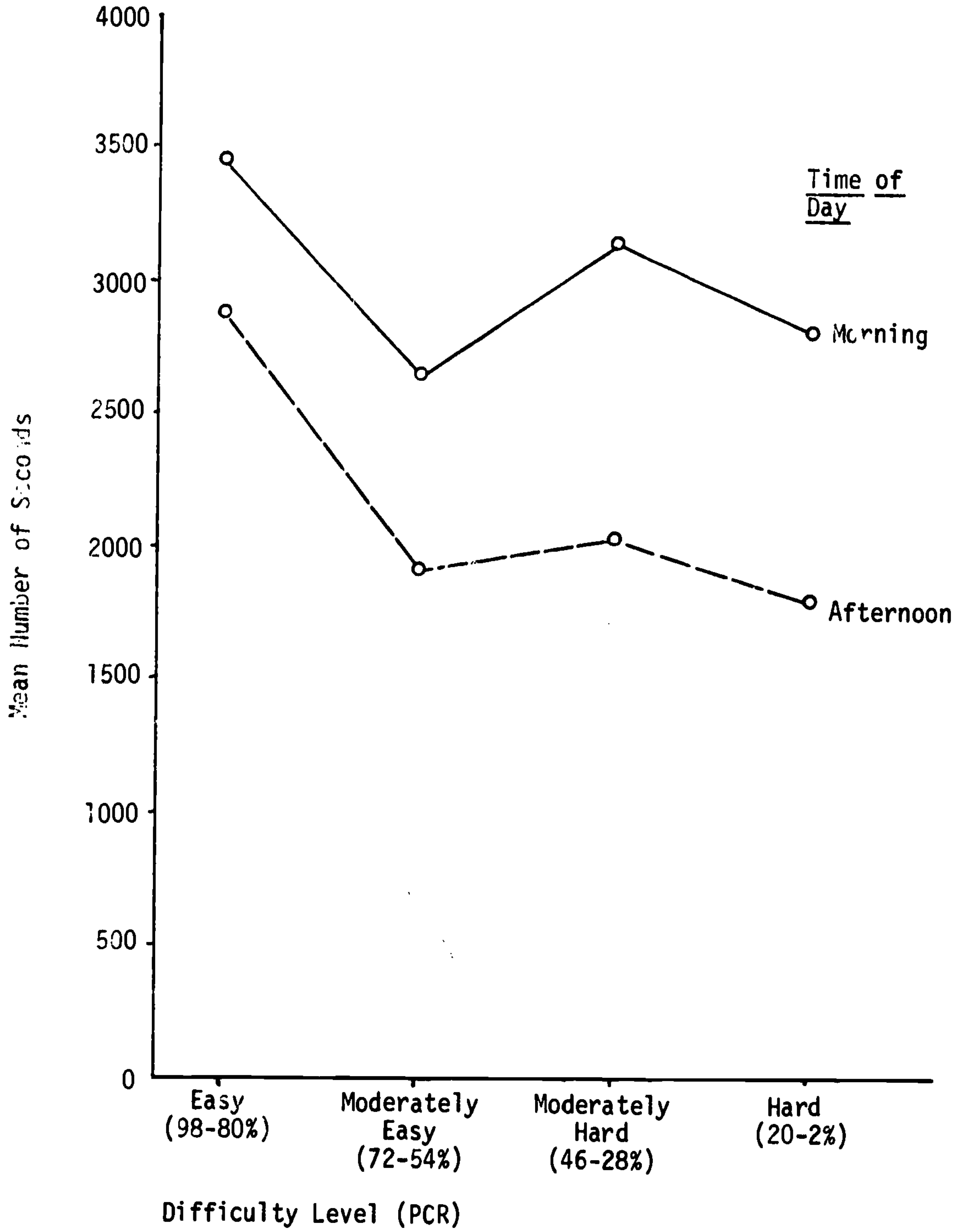


Fig. 26. Interaction of Time of Day with Task Difficulty (PCR) with Task Time as the Dependent Variable, One Aspect of the Three Way Interaction between Time of Day, Incentive, and Task Difficulty.

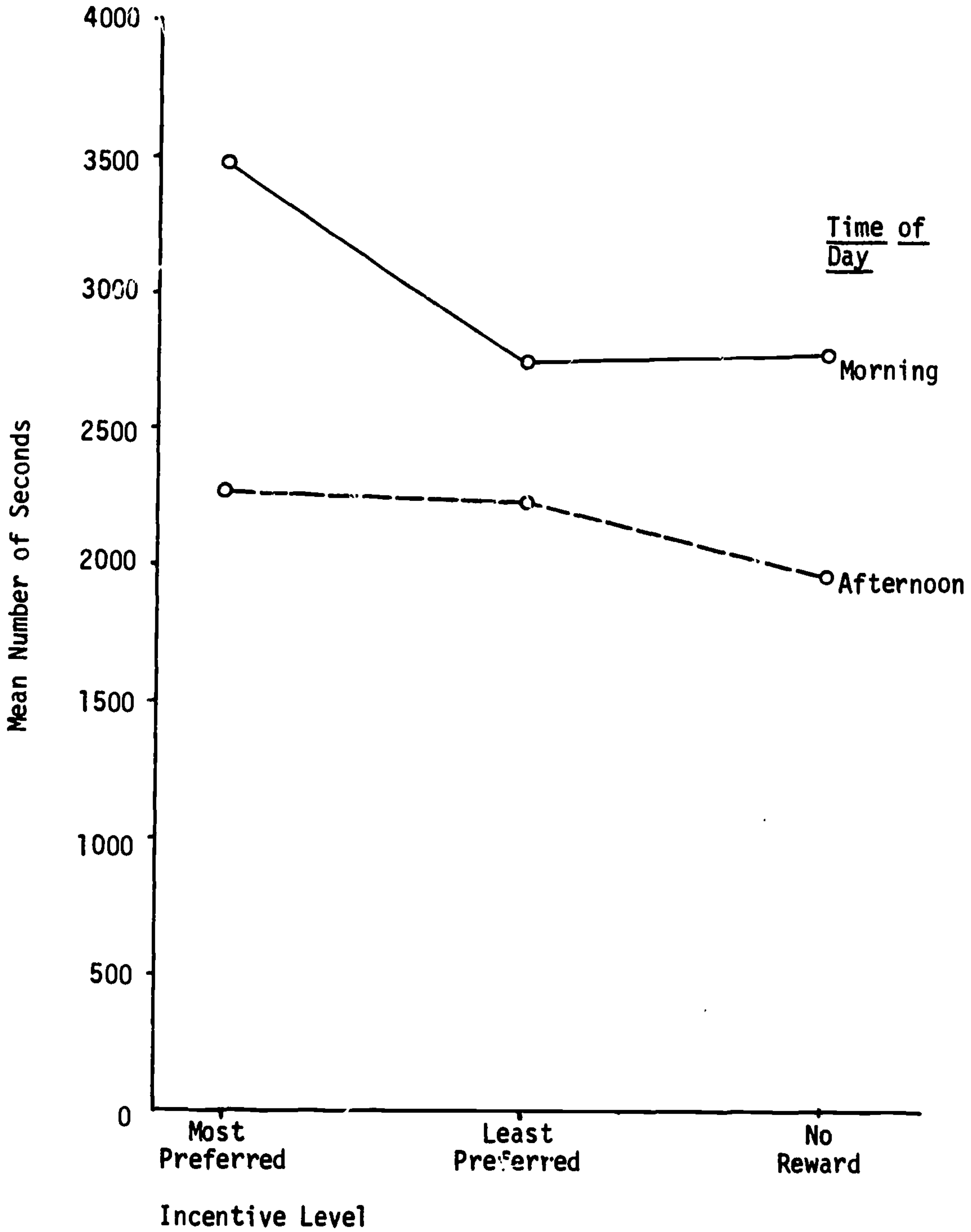


Fig. 27. Interaction of Time of Day with Incentive with Task Time as the Dependent Variable, One Aspect of the Three Way Interaction between Time of Day, Incentive, and Task Difficulty.

Table 29

Analysis of Variance Summary Table for Incentive, PCR, and Time of Day, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive	4300.44	2	2150.22	<1.0	0.589
2 PCR	358219.27	3	119406.42	29.515	0.000
3 Time of Day	12018.99	1	12018.99	2.971	0.088
12	21698.03	6	3616.34	<1.0	0.502
13	909.12	2	454.56	<1.0	0.894
23	59.37	3	19.79	<1.0	0.999
123	8016.23	6	1336.04	<1.0	0.920
Error	449059.39	111	4045.58		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 64.19 with 23 d.f., $p<.001$. The Hypothesis of Homogeneity of Variance is rejected.

Assuming proportional population subclass frequencies, the probability of obtaining a sample with the obtained subclass frequencies is less than 0.1.

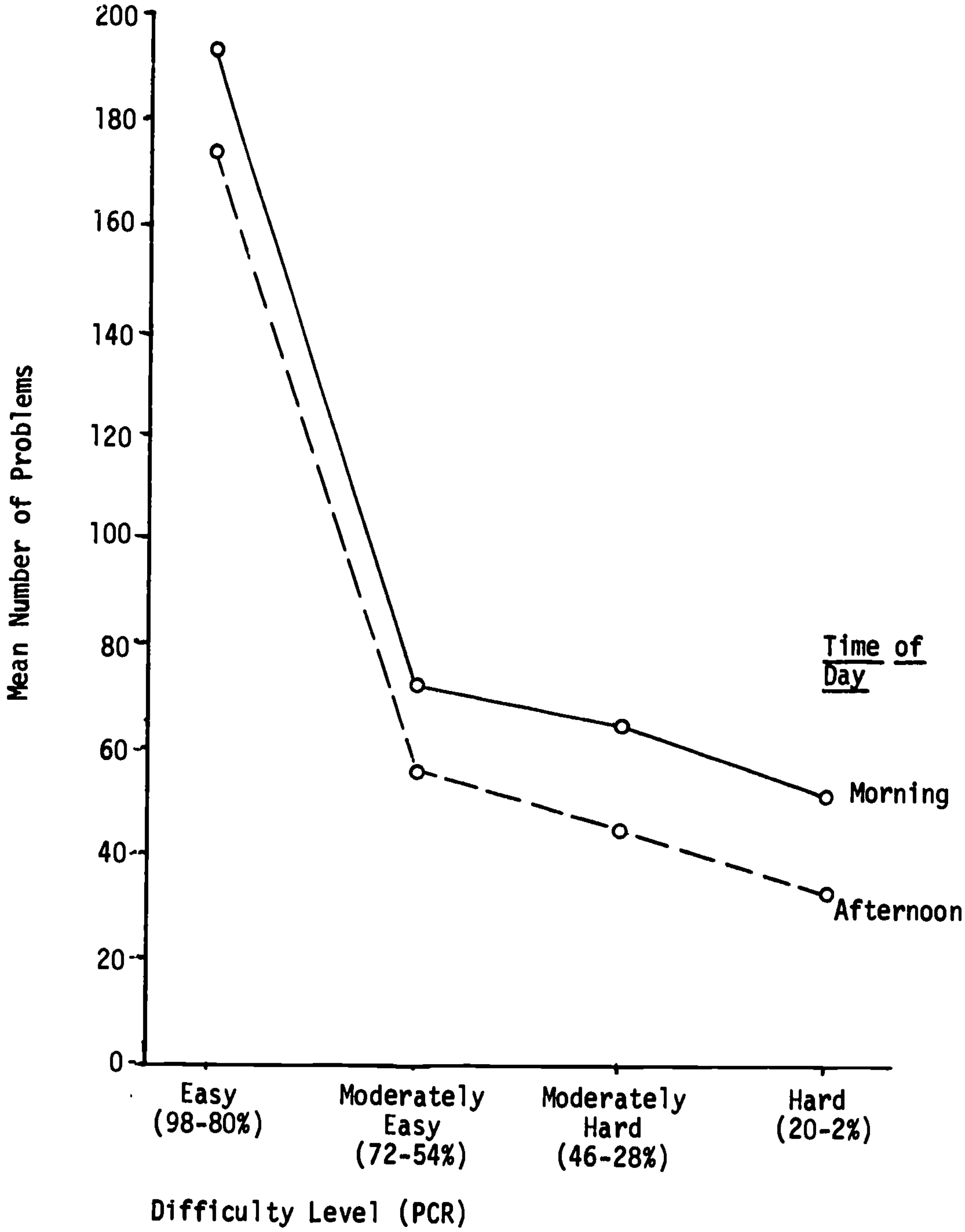


Fig. 28. Interaction between Time of Day and Task Difficulty (PCR) with ANP as the Dependent Variable, One Aspect of the Three Way Interaction between Time of Day, Incentive, and Task Difficulty.

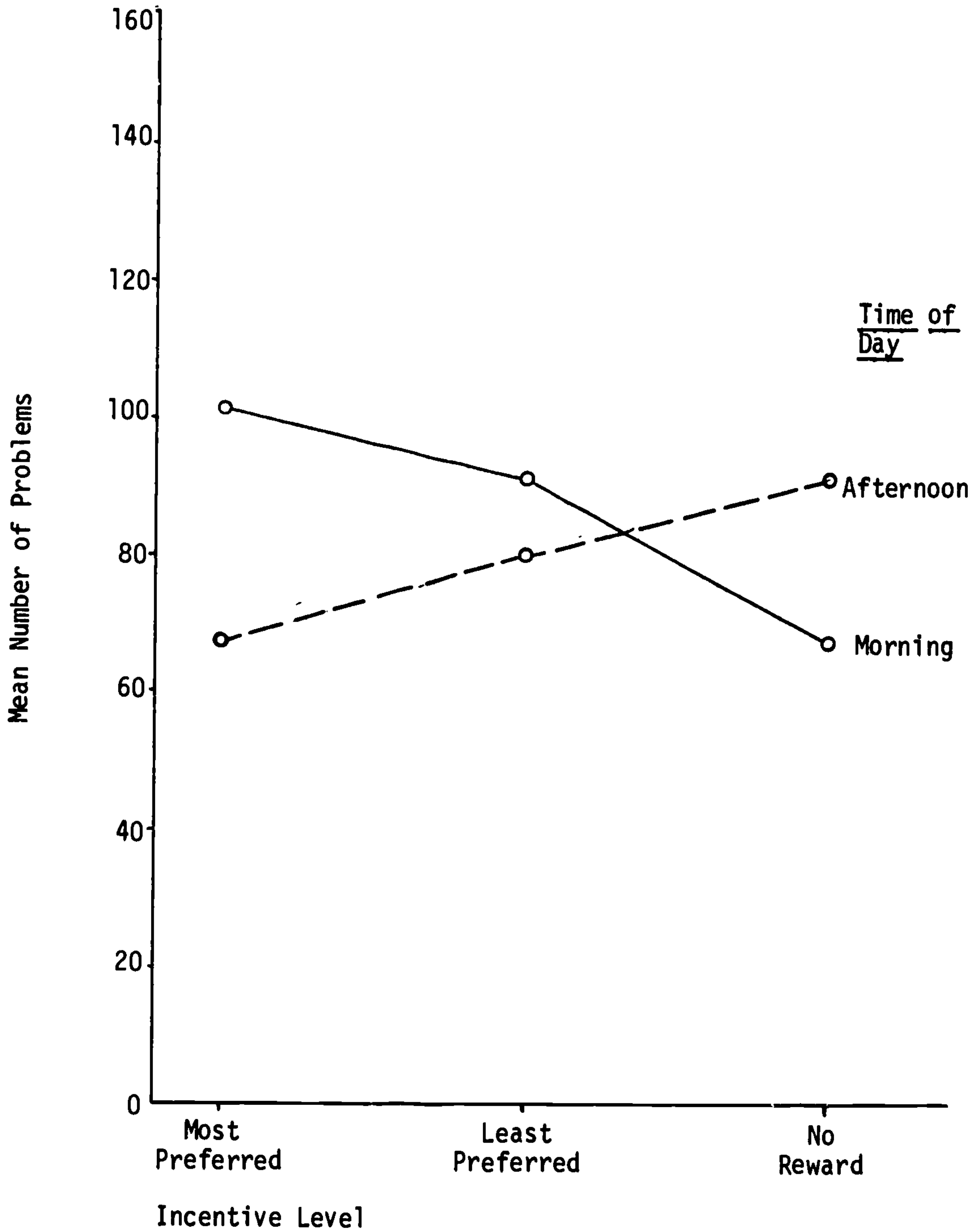


Fig. 29. Interaction between Time of Day and Incentive with ANP as the Dependent Variable, One Aspect of the Three Way Interaction between Time of Day, Incentive, and Task Difficulty.

Table 30

Analysis of Variance Summary Table for Incentive Preference and Incentive Kind, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive Pref.	5768583.	1	5768583.	2.485	0.119
2 Incentive Kind	11059934.	2	5529967.	2.382	0.099
12 Pref. x Kind	11736371.	2	5868185.	2.528	0.086
Error	192703281.	83	2321726.		

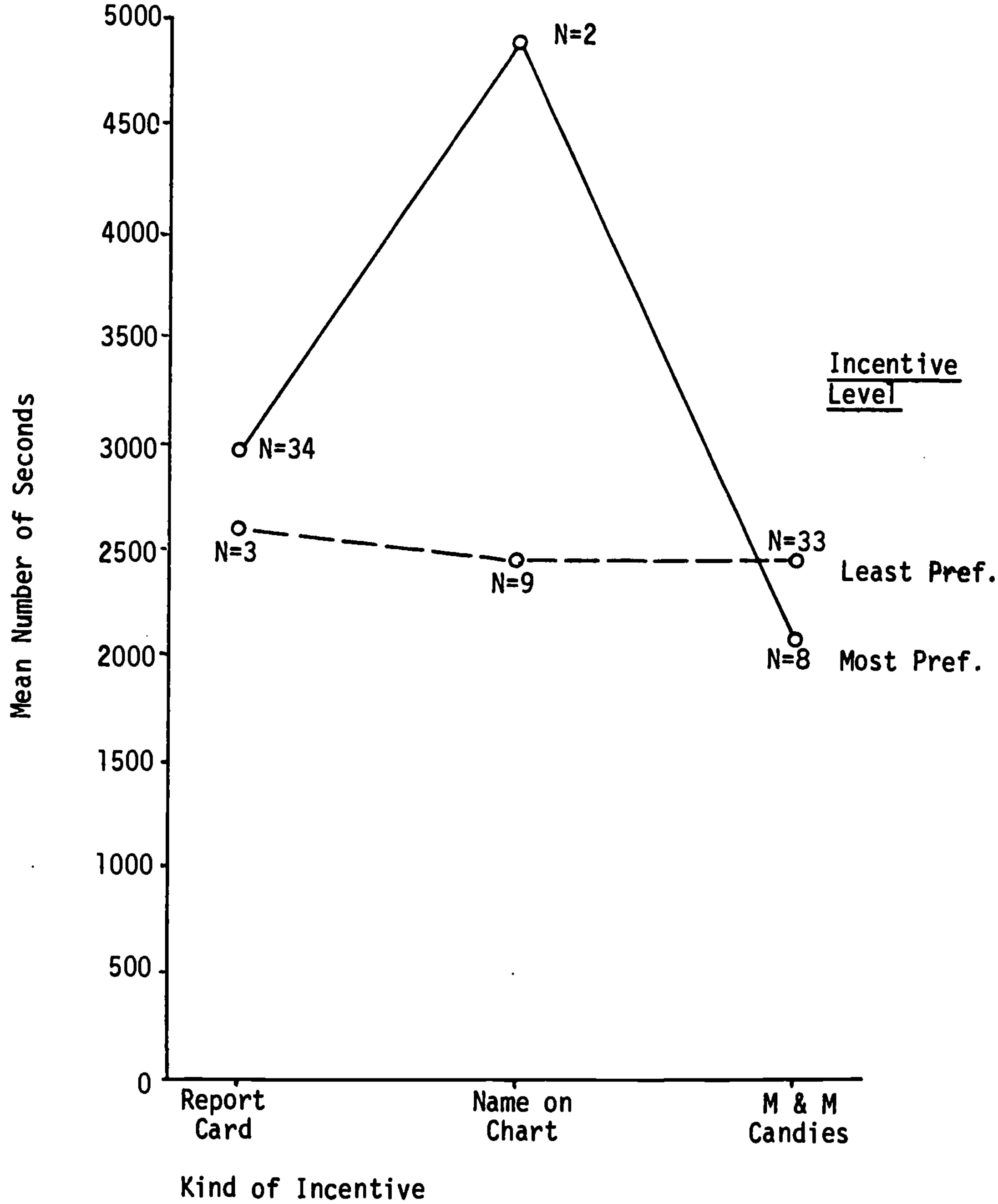


Fig. 30. Interaction between Incentive Level and Kind of Incentive with Task Time as the Dependent Variable.

ERIC

Table 31

Analysis of Variance Summary Table for Incentive Preference and Incentive Kind, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Incentive Pref.	935.3	1	935.3	<1.0	0.712
2 Incentive Kind	27552.6	2	13776.3	2.018	0.139
12 Pref. x Kind	3647.6	2	1823.8	<1.0	0.766
Error	566730.0	83	6828.1		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 15.71 with 5 d.f., $p<.008$. The Hypothesis of Homogeneity of Variance is rejected.

Assuming proportional population subclass frequencies, the probability of obtaining a sample with the obtained subclass frequencies is less than 0.1.

ERIC Full Text Provided by ERIC

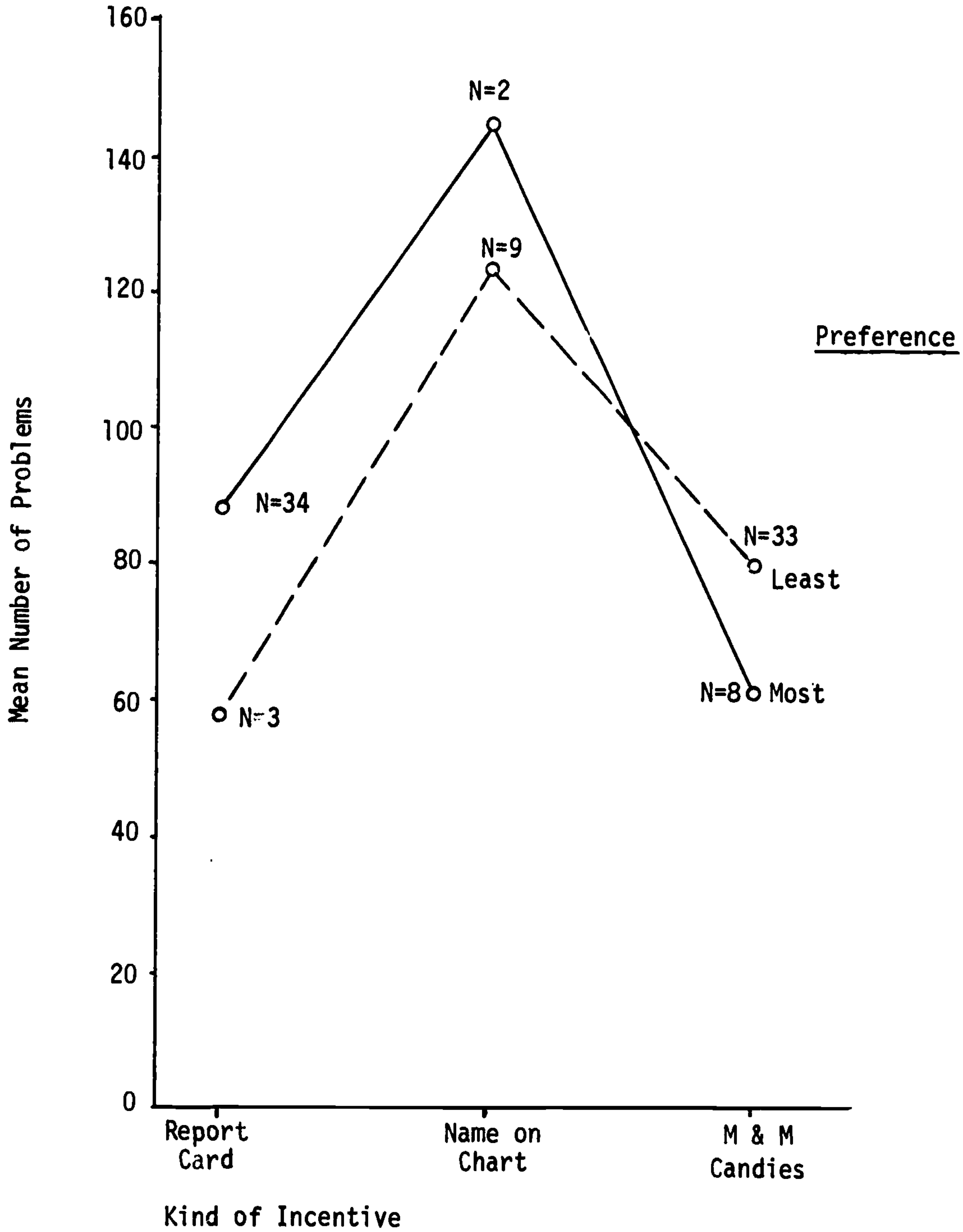


Fig. 31. Interaction between Incentive Level and Kind of Incentive with ANP as the Dependent Variable.

Table 32

Analysis of Variance Summary Table for Incentive Kind and Time of Day, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Time of Day	14890488.	1	14890488.	5.566	0.020
2 Incentive Kind	5133016.	3	1711005.	<1.0	0.591
12 TOD x Kind	3456328.	3	1152109.	<1.0	0.731
Error	339736554.	127	2675091.		

Note: For Bartlett's Test of Homogeneity of Variance, Chi-Square = 14.57 with 7 d.f., $p = .042$. The Hypothesis of Homogeneity of Variance is rejected.

ERIC

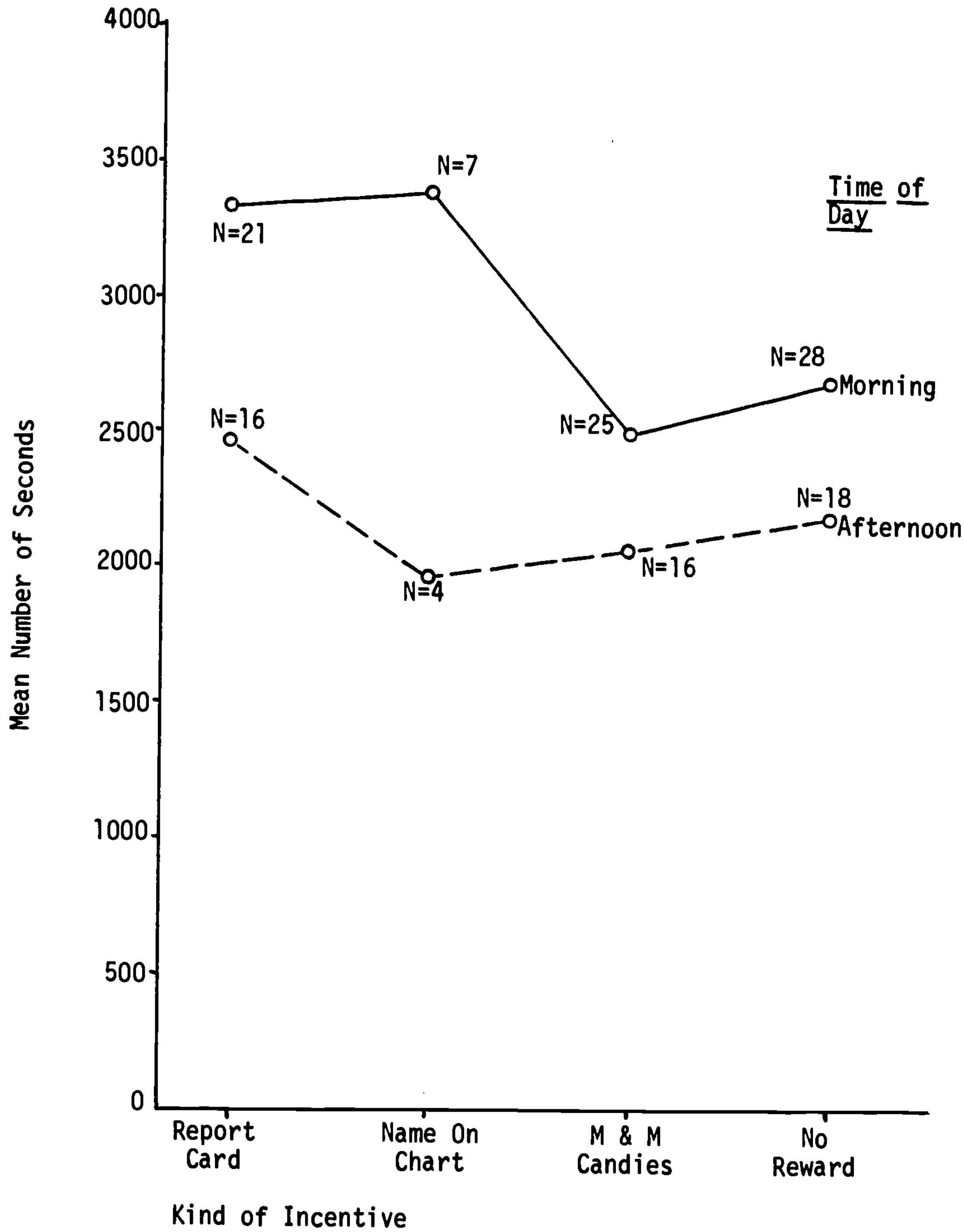


Fig. 32. Interaction between Time of Day and Kind of Incentive with Task Time as the Dependent Variable.

Table 33

Analysis of Variance Summary Table for Incentive Kind and Time of Day, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Time of Day	13228.	1	13228.	2.043	0.155
2 Incentive Kind	21072.	3	7024.	1.085	0.358
12 TOD x Kind	46596.	3	15532.	2.399	0.071
Error	822331.	127	6475.		

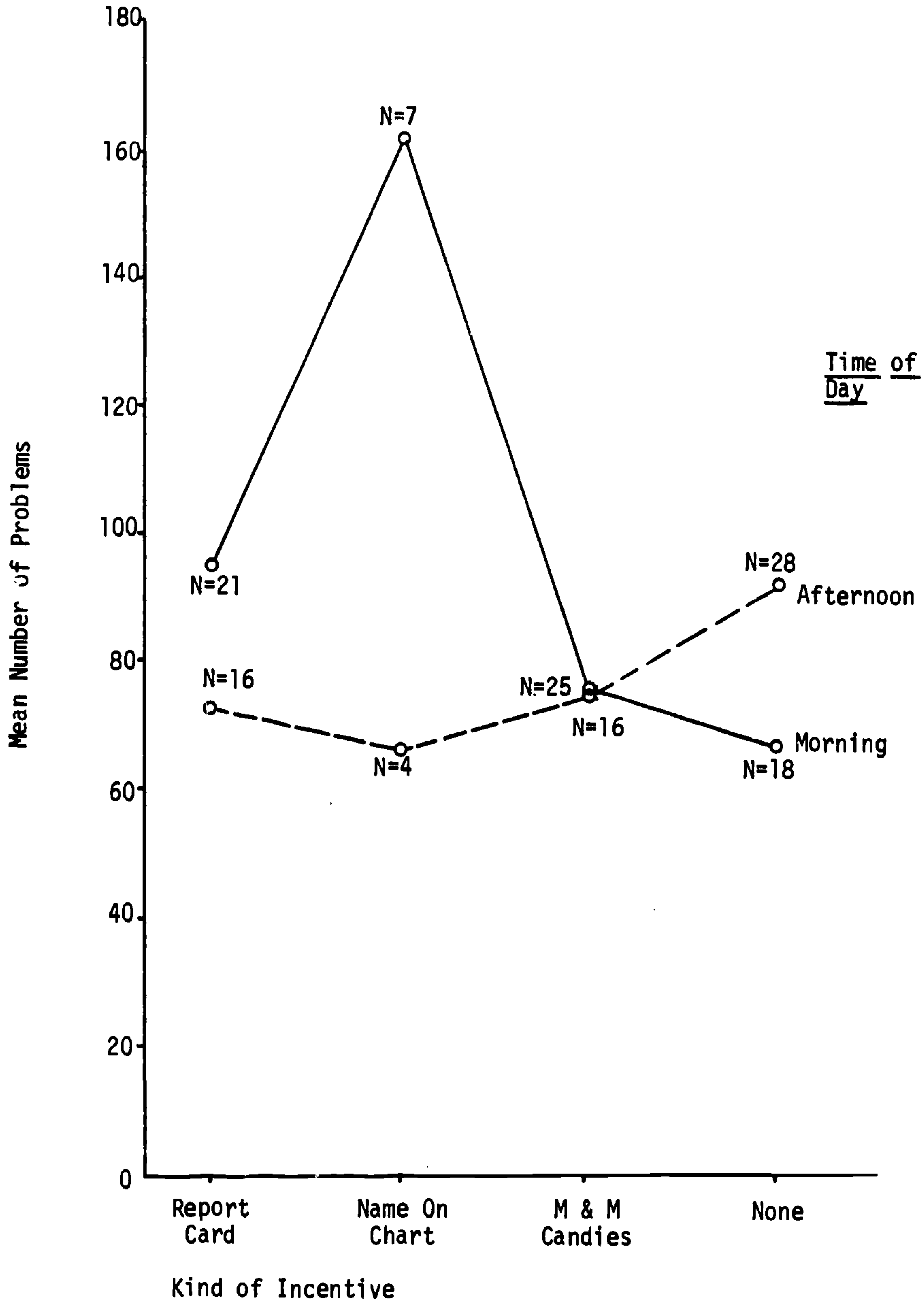


Fig. 33. Interaction of Time of Day and Kind of Incentive with ANP as the Dependent Variable.

Table 34

Analysis of Variance Summary Table for Incentive Kind and Sex, all Subjects, with Task Time as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Sex	10562289.	1	10562289.	3.922	0.050
2 Incentive Kind	6092680.	3	2030893.	<1.0	0.522
12 Sex x Kind	3640134.	3	1213378.	<1.0	0.717
Error	342022584.	127	2693091.		

ERIC

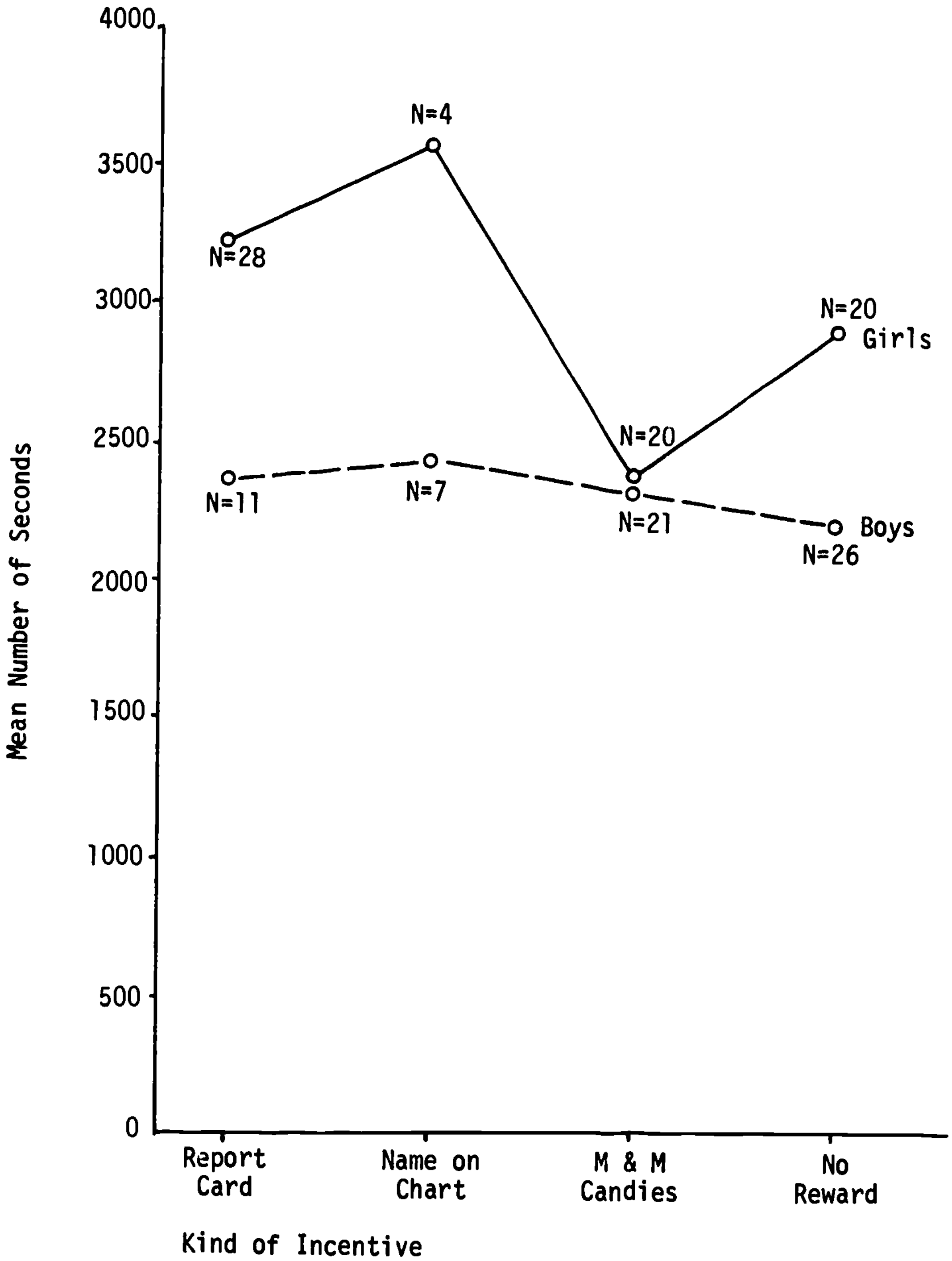


Fig. 34. Interaction between Sex and Kind of Incentive with Task Time as the Dependent Variable.

Table 35

Analysis of Variance Summary Table for Incentive Kind and Sex, all Subjects, with ANP as the Dependent Variable

Source	Sums of Squares	DF	Mean Squares	F Ratio	Probability
1 Sex	20870.	1	20870.	3.178	0.077
2 Incentive Kind	61815.	3	20605.	3.138	0.028
12 Sex x Kind	22490.	3	7497.	1.142	0.335
Error	833943.	127	6566.		

Note: Assuming proportional population subclass frequencies, the probability of obtaining a sample with the obtained subclass frequencies is less than 0.1.

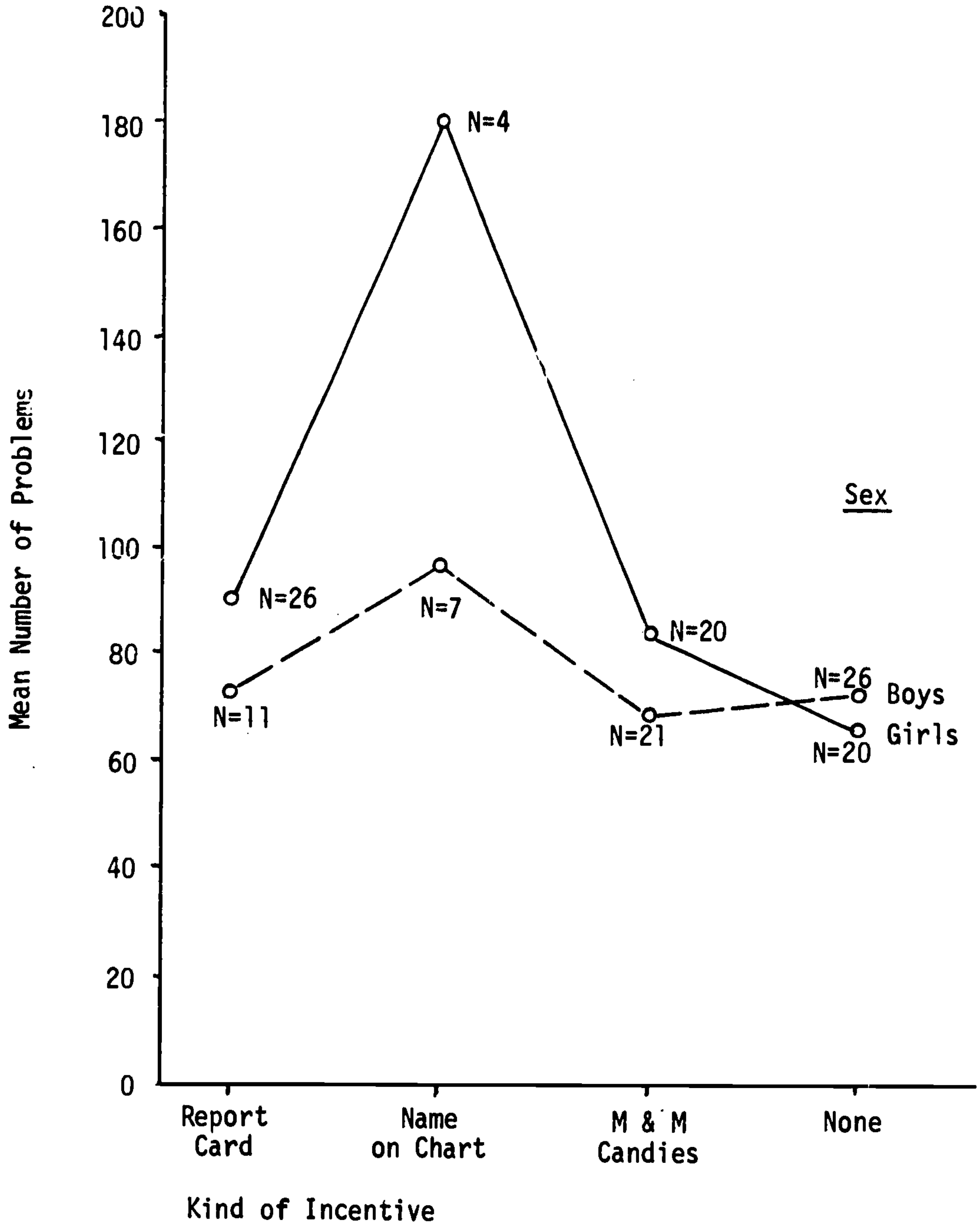


Fig. 35. Interaction between Sex and Kind of Incentive with ANP as the Dependent Variable.

APPENDIX L

Data Pertaining to the
Medians Analysis

ERIC

Table 36

Cell and Marginal Medians and Means for the Results for the Task Time Variable, in Seconds

Incentive Value	Easy (98-80%)	Mod. Easy (72-54%)	Mod. Hard (46.28%)	Hard (20-2%)	Marginal Data
Most Preferred	N = 10 Mdn = 3087.3 (Mean = 3320)	N = 13 Mdn = 2282.4 (Mean = 2943)	N = 11 Mdn = 2146.8 (Mean = 2720)	N = 10 Mdn = 2558.5 (Mean = 2666)	N = 44 Mdn = 2604.4 (Mean = 2910)
Least Preferred	N = 9 Mdn = 2606.7 (Mean = 3344)	N = 14 Mdn = 1583.5 (Mean = 2092)	N = 8 Mdn = 3335.2 (Mean = 2910)	N = 14 Mdn = 1685.4 (Mean = 1837)	N = 45 Mdn = 1946.8 (Mean = 2409)
No Reward	N = 10 Mdn = 3122.7 (Mean = 3188)	N = 15 Mdn = 1030.4 (Mean = 2050)	N = 11 Mdn = 1868.0 (Mean = 2261)	N = 10 Mdn = 1831.6 (Mean = 2699)	N = 46 Mdn = 2459.3 (Mean = 2489)

Table 37

Cell and Marginal Medians and Means for the Results for the Adjusted Number of Problems Variable

Incentive Value	Easy (98-80%)	Mod. Easy (72-54%)	Mod. Hard (46-28%)	Hard (20-2%)	Marginal Data
Most Preferred	N = 10 Mdn = 143.7 (Mean = 158.0)	N = 13 Mdn = 88.0 (Mean = 79.9)	N = 11 Mdn = 56.6 (Mean = 58.4)	N = 10 Mdn = 48.5 (Mean = 52.2)	N = 44 Mdn = 79.0 (Mean = 86.0)
Least Preferred	N = 9 Mdn = 167.0 (Mean = 219.8)	N = 14 Mdn = 42.5 (Mean = 65.1)	N = 8 Mdn = 59.5 (Mean = 60.6)	N = 14 Mdn = 26.5 (Mean = 38.6)	N = 45 Mdn = 49.0 (Mean = 87.0)
No Reward	N = 10 Mdn = 170.0 (Mean = 177.7)	N = 15 Mdn = 40.5 (Mean = 53.7)	N = 11 Mdn = 35.0 (Mean = 52.0)	N = 10 Mdn = 31.0 (Mean = 38.4)	N = 46 Mdn = 55.0 (Mean = 76.9)

Table 38

Division of the Number of Observations
Above and Below the Overall Median
of 2423.4 for Task Time Data

Incentive Value	Location	Easy (98-80%)	Mod. Easy (72-54%)	Mod. Hard (46-28%)	Hard (20-2%)	Marginal Data
Most Preferred	Above	7	6	5	5	23
	Below	3	7	6	5	21
Least Preferred	Above	6	3.5	5	5	19.5
	Below	3	10.5	3	9	25.5
No Reward	Above	9	6	4	5	24
	Below	1	9	7	5	22
Marginal Data	Above	22	15.5	14	15	
	Below	7	26.5	16	19	

Table 39

Division of the Number of Observations Above and Below the Overall Median of 57.8 for Adjusted Number of Problems Data

Incentive Value	Location	Easy (98-80%)	Mod. Easy (72-54%)	Mod. Hard (46-28%)	Hard (20-2%)	Marginal Data
Most Preferred	Above	10	8	4	4	26
	Below	0	5	7	6	18
Least Preferred	Above	9	3	5	3	20
	Below	0	11	3	11	25
No Reward	Above	8	6	4	4	22
	Below	2	9	7	6	24
Marginal Data	Above	27	17	13	11	
	Below	2	25	17	23	

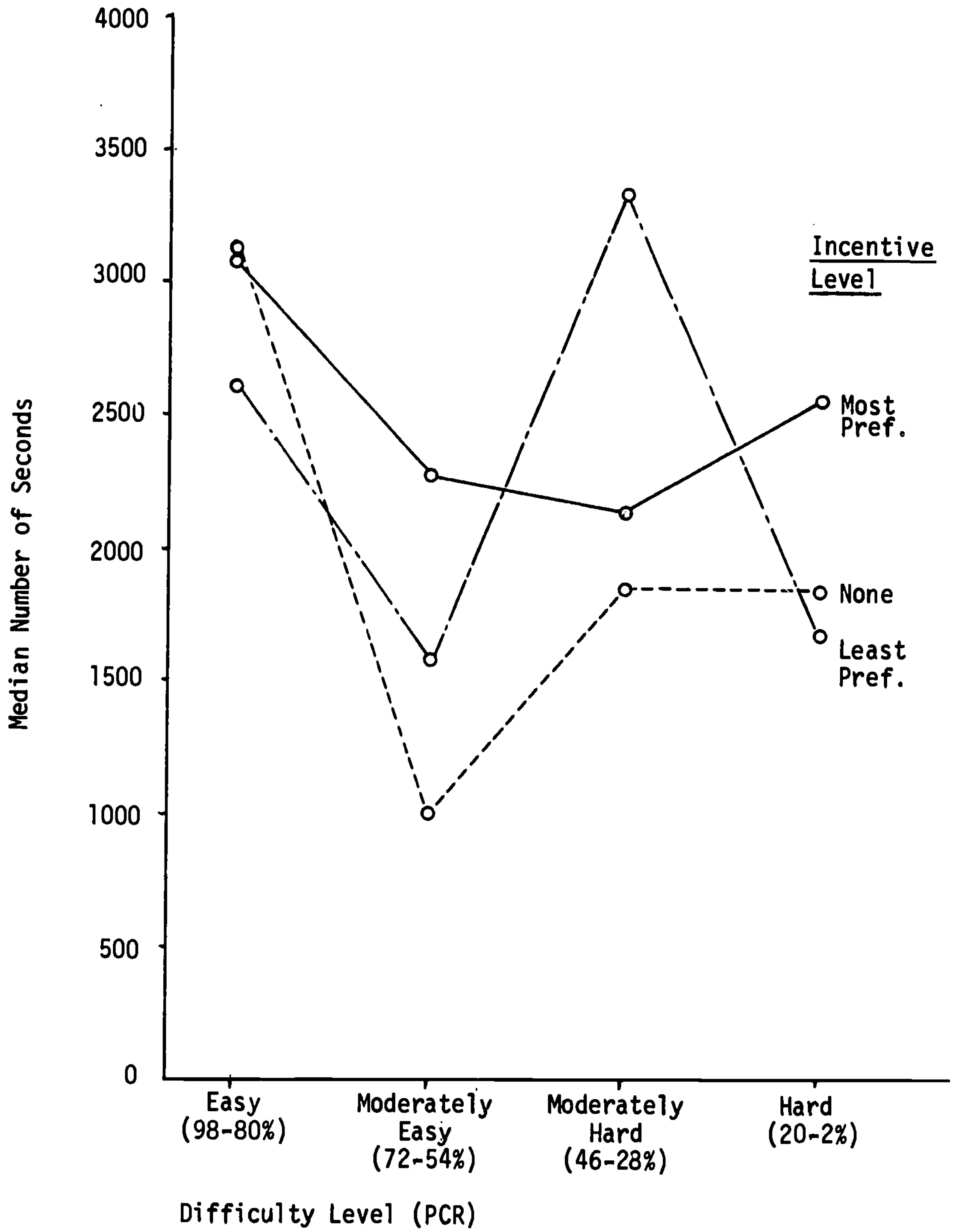


Fig. 36. Plot of Incentive against Task Difficulty (PCR) with Median Task Time (in Seconds) as the Dependent Variable.

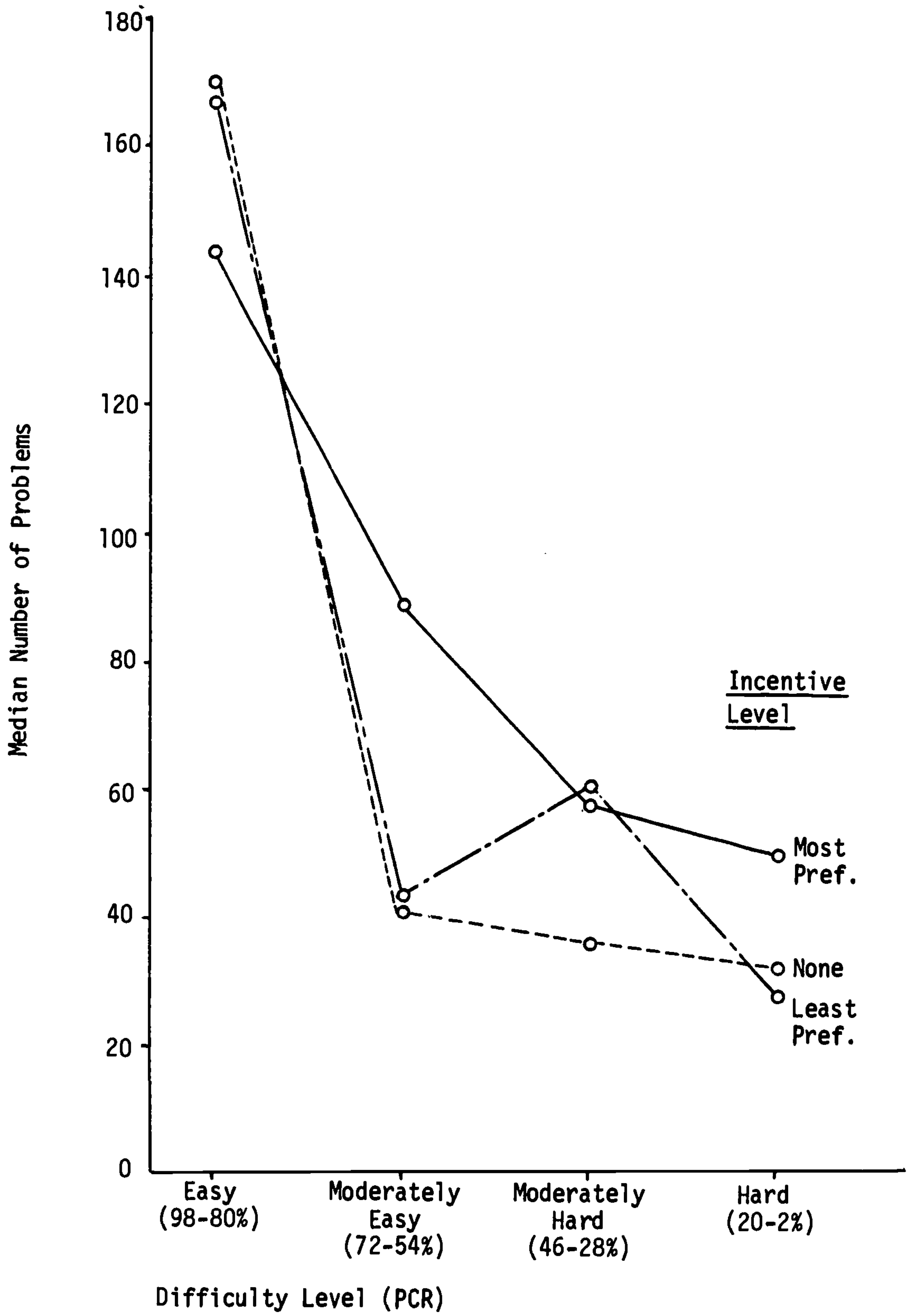


Fig. 37. Plot of Incentive against Task Difficulty (PCR) with Median Adjusted Number of Problems as the Dependent Variable.